REPORT RESUMES

ED 002 819 24

A STUDY OF ESTHETIC JUDGMENT.

BY- CHILD, IRVIN L.
YALE UNIV., NEW HAVEN, CONN.

REPORT NUMBER CRP-669 PUB DATE 62
EDRS PRICE MF-$0.27 HC-$5.28 132P. JOURNAL OF
PERSONALITY, 30/496-512/1962

DESCRIPTORS- *CULTURAL ENRICHMENT, *ART APPRECIATION, *COLLEGE STUDENTS, *DISCRIMINATORY ATTITUDES (SOCIAL), PAINTING, PERSONALITY, GROUP DYNAMICS, NEW HAVEN, CONNECTICUT, STANFORD, CALIFORNIA

THE ABILITY OF COLLEGE STUDENTS TO RECOGNIZE DEGREES OF ESTHETIC MERIT IN OBJECTS OF ART WAS STUDIED. THE OBJECTIVE WAS TO DETERMINE BY EXPERIMENTATION SOME OF THE FACTORS WHICH MAY BE IMPORTANT IN THE DEVELOPMENT OF ESTHETIC JUDGMENT. A SAMPLE OF MALE COLLEGE STUDENTS WAS GIVEN SEVERAL TESTS OF ESTHETIC JUDGMENT. FROM THESE SUBJECTS WERE CHOSEN SETS OF SUBJECTS, EQUATED ON JUDGMENT OF ART. EACH SET WAS DIVIDED AT RANDOM INTO VARIOUS EXPERIMENTAL GROUPS. THE GROUPS WERE REQUIRED TO RESPOND WITH ESTHETIC JUDGMENT TO A LARGE NUMBER OF PAIRS OF ART OBJECTS PRESENTED ON SLIDES. ESTHETIC JUDGMENT WAS FOUND TO BE RELATED TO AMOUNT OF BACKGROUND IN ART. THERE WAS NO WAY OF ASSESSING TO WHAT EXTENT THIS REFLECTS AN INFLUENCE OF BACKGROUND ON ESTHETIC JUDGMENT SCORE AND TO WHAT EXTENT IT REFLECTS A TENDENCY FOR PEOPLE WITH BETTER ESTHETIC JUDGMENT TO BE MORE INTERESTED IN ART AND THUS OBTAIN MORE EXPERIENCE WITH IT. THE RESULTS OF THIS STUDY CONSTITUTE TENTATIVE FINDINGS WHICH MIGHT WELL SERVE AS SOURCES OF HYPOTHESES IN THE STUDY OF OTHER GROUPS. OF SPECIAL IMPORTANCE WAS THE GENERAL PATTERN INTO WHICH A NUMBER OF FINDINGS FELL, SUGGESTING THAT GOOD ESTHETIC JUDGMENT IS LARGELY AN OUTCOME OF A GENERAL COGNITIVE APPROACH TO THE WORLD, AN APPROACH INVOLVING SEARCH FOR COMPLEX AND NOVEL EXPERIENCE WHICH IS THEN UNDERSTOOD AND EVALUATED THROUGH RELATIVELY AUTONOMOUS INTERACTIONS OF THE INDIVIDUAL WITH OBJECTS WHICH PROVIDE SUCH EXPERIENCE. (JH)

ED 002819

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

A STUDY OF ESTHETIC JUDGMENT

Cooperative Research Project No. 669

Irvin L. Child

Yale University

New Haven, Connecticut

1962

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

CONTENTS

Page No.

CHAPTER I. - Personal Preferences as an Expression of Esthetic Sensitivity 1

CHAPTER II. - Observations on the Meaning of some Measures of Esthetic Sensitivity. 18

CHAPTER III. - Personality Correlates of Esthetic Judgment 34

CHAPTER IV. - Experiments on Learning to Discriminate between Better and Poorer Works of Art . . 84

APPENDIX 115

TABLES AND ILLUSTRATIONS

Page No.

CHAPTER I -

Table 1. Degree of agreement among persons in personal preferences and esthetic judgments 5

Table 2. Correlation between ordering of pictures by group preference and by esthetic value 8

Table 3. Agreement of individual preferences with group preference and with esthetic value: relation of these measures to other variables 11

CHAPTER II -

Table 1. Correlation of Bulley and Graves tests with other variables 23

Figure 1. Split-half reliability of two-choice items in Graves Design Judgment Test. 25

Table 2. Intercorrelations of variables for 22 subjects in preference study . . . 30

Chapter III -

Table 1. Intercorrelations of background variables, and relation to esthetic judgment 38

Table 2. Reliability and intercorrelations of tolerance measures 54

Chapter I

PERSONAL PREFERENCES AS AN EXPRESSION OF ESTHETIC SENSITIVITY[1]

In psychological research on esthetics, people have often been asked to rate or rank a set of stimuli according to personal liking. A set of color patches may be presented, for example, and the S asked to rank them in order of preference (Ludvigh, as cited by Beebe-Center, 1932, p. 208; H. J. Eysenck, 1941; Granger, 1955). Or the same may be done with a set of odorants (Beebe-Center, 1932, p. 203; M. D. Eysenck, 1944), or with a set of paintings or other works of art (H. J. Eysenck, 1940).

When this is done, there always seems to be some tendency toward agreement among different persons in their order of preference. What then is the significance of the average or group order of preference which emerges? It has been frequently taken to represent an ordering of the stimuli according to their relative esthetic value. Sometimes this suggestion has been put forth simply as a tentative operational definition of esthetic value (as it was by Beebe-Center, 1932, pp. 200-202); sometimes it has been accompanied by the assertion of a testable relation to some other criterion of esthetic value (as, for example, by H. J. Eysenck, 1957, pp. 329-330). An argument presented in an effort to make such a definition or assertion reasonable is that the idiosyncratic preferences of different people will cancel each other out, leaving only the consistent, panhuman regularities of esthetic appeal as a basis for agreement in preference (Beebe-Center, 1932, p. 202; Granger, 1955, p. 17; H. J. Eysenck, 1957, p. 324).

In research of this kind, it is also found that people differ in the extent to which their individual preferences agree with the average order of preference. What is the meaning of this kind of variation among persons? It has often been assumed that it provides a measure of esthetic sensitivity. The person whose degree of liking for various stimuli corresponds fairly closely to the group average is said to be highly sensitive to esthetic values, either in general or for the particular kind of stimuli being used; or to have good taste. The person whose preferences are little related to the consensus, or even tend to be contrary to it, is said to be low in esthetic sensitivity, or in goodness of taste. Again, such a statement seems to be offered sometimes merely as a definition (Beebe-Center, 1932, p. 203), sometimes also as a testable hypothesis (H. J. Eysenck, 1957, pp. 319-320).

---

1. Part of the research reported in this chapter was done with financial assistance from the Center for Advanced Study in the Behavioral Sciences, while the author was a fellow there, and was facilitated by the very generous provision of facilities by the Department of Psychology at Stanford University. In nearly the present form, this chapter has been published in The Journal of Personality, 1962, volume 30, pp. 496-512.

Such statements, when offered as definitions of esthetic value of stimuli or of the esthetic sensitivity of persons, seem to have little in common with the ordinary meaning of these terms, as is indeed recognized by H. J. Eysenck (1957, p. 323). A body of very interesting research has, however, been reported in the context of such statements. Major examples are presented or reviewed in the references given above. Some of this research does seem to establish an empirical relation between esthetic value or esthetic sensitivity, defined in this unusual way, and more usual meanings of these terms. H. J. Eysenck (1957) has even been led to base a general account of "The Psychology of Esthetics" primarily on research using this approach of popular-preference consensus.

This research merits careful scrutiny and further pursuit. Long ago Stephenson (1936) objected to the unreasonableness of the definitions involved, and even the most recent review of this research expresses amazement at some of the empirical findings (Valentine, 1962, p. 92). Such responses to this research are highly appropriate but are not in themselves sufficient, for if the findings are valid they are of fundamental importance for the psychology of esthetics and of personality.

In order to place this research in proper perspective, I believe the essential first step is to apply the method to stimuli for which an external criterion of esthetic value is available, and to relate findings to that criterion. This implies a deliberate return to the kind of stimuli -- works of art varying in esthetic value according to ordinary criteria -- used in the pioneering work of Burt (1933, and briefly described in a more accessible publication by Pickford, 1955). Such stimuli are called for even though, as H. J. Eysenck (1940) has pointed out, reactions to them are much more complicated than are reactions to single colors, odorants, etc., for which, unhappily, an external criterion of esthetic value is not readily available and perhaps would have no real significance. It would also be useful, in evaluating the meaning of individual differences, to have an entirely independent measure of each person's esthetic sensitivity in a more ordinary sense. In earlier research, this need seems to have been recognized only by Burt (1933) and by Granger (1955).

In the present research, then, the following questions will be dealt with in order: if Ss unselected for competence in art are asked to rate works of art according to personal preference, will there be a consensus of preference? Is degree of agreement with the group rating a reliable characteristic of the individual S? Does the evaluation of works of art by average preference rating agree with their esthetic value as determined by a more self-evident criterion? Is the degree to which an individual's preferences agree with esthetic value, as determined by this more obvious criterion, a reliable characteristic of the individual? Are these two characteristics of the individual -- agreement with average rating and agreement with esthetic value -- correlated with each other? Finally, what is the relation of each of these two measures to other personal characteristics, including an independent measure of esthetic sensitivity?

METHOD

Stimulus Materials

The basic data of this research consisted of ratings made by paid Ss on a large number of postcard-size reproductions of paintings. There were 12 sets, each comprising 60 pictures. The sets, in the order of their presentation, were as follows:

1. Pictures of groups of people
2. Religious pictures
3. Landscapes
4. Abstract (i.e., nonrepresentational) pictures
5. Pictures of single men
6. A second set of pictures of groups of people
7. A second set of religious pictures
8. Pictures which markedly distort visual reality, or which represent fantasy
9. Pictures of single women
10. Pictures of children (single or in groups composed only of children)
11. Pictures of animals, or including animals
12. Still lifes

Sets 1 and 6 were obtained by dividing 120 pictures of groups of people at random into two halves. Sets 2 and 7, of religious pictures, were obtained in a parallel way. Within each of the 12 sets, an attempt was made at great diversity of subject matter, period, style, and probable appeal, within the limits set by the availability of postcards at book stores and art-supply stores.

Subjects

Two sets of male college students were used as Ss in a study of preferences. The first set of 22 Ss all were sophomores at Stanford University, mostly secured through the introductory psychology course but a few through the student employment bureau; they all were from the western states. The second set of 22 Ss were undergraduates at Yale University, all obtained through a second-semester course in psychology; they were from all regions of the United States, though predominantly from the Northeast. The Stanford study was conducted by the writer, the Yale study by a student assistant with no previous experience in psychological research. Most Ss appear to have volunteered in order to earn money, and all appeared to carry out seriously the responsibility they had assumed. In general, Ss were not allowed to know in advance that the research was concerned with art, in order that they would not be selected with respect to interest or competence in art. At Yale there were probably no exceptions; at Stanford there may have been one or two.

Procedure

The data were collected entirely in individual sessions. The S was asked to react to each set of pictures by dividing them into

10 piles of 6 each, according to the appeal of each picture to him, from the 6 he liked best to the 6 he liked least. The few Ss who asked whether they should use various other specific criteria were told that they should concentrate on how well they liked or disliked the pictures, and that other criteria were relevant only insofar as they influenced this one. Most Ss spontaneously finished with each set within 15 or 20 minutes. The few who took longer were encouraged to go faster if they found they reasonably could, but were not prevented from continuing to take more time if they felt they needed it.

After reacting to all the sets of pictures, Ss were interviewed about the bases for their preferences and about their life history in relation to art. They were then, or at a later session, given several personality tests; those relevant to the present report will be mentioned later.

Criterion Judgments

To obtain a criterion of esthetic value for the pictures, an entirely different task was set for a third group of people. They were asked to judge the esthetic merit of the pictures (i.e., how good they are as works of art), by dividing each set of pictures into 10 piles of approximately 6 each. The persons used for this task were of both sexes and were mostly graduate students of art or of the history of art, but included several other persons who felt able to make the judgments required. As the purpose here was to get a measure of average opinion on a matter on which there was generally rather good agreement, not to study individual differences, only 13 judges were used. They generally required two or three times as long as the students who were expressing preferences, though some did not.

RESULTS

Agreement in Preference Orders

A first question that may be asked is this: When paintings of great diversity are rated for personal preference by college students not selected for interest or competence in art, is there any agreement among the students in their order of preference? This may be tested first of all by calculating, for each set of pictures, a coefficient of correlation as a measure of the degree of similarity between the ordering by each person and by every one of the other 21 persons in the group. For each set, there are both positive and negative correlations (the latter indicating disagreement), but the median correlation is always positive in both groups of Ss, as shown in columns A and B of Table 1.

A second way of testing this general question is of special interest because it enables us to be more specific, and to test in a simple manner whether a tendency toward agreement is present in all, most, or merely some smaller fraction of the Ss. This is done by determining what the group ordering of the picture is, and measuring the

TABLE 1

DEGREE OF AGREEMENT AMONG PERSONS IN PERSONAL PREFERENCES AND ESTHETIC JUDGMENTS

Set of pictures	Median interperson correlation			Agreement between group preferences of Stanford and Yale groups (D)	Number of negative saturations on first factor		
	Expressions of preference		Judgments of esthetic value (C)		Expressions of preference		Judgments of esthetic value (G)
	Stanford (A)	Yale (B)			Stanford (E)	Yale (F)	
1. Groups I	+.21	+.18	+.53	+.91	1	0	0
2. Religious I	+.28	+.15	+.40	+.84	1	0	0
3. Landscapes	+.29	+.23	+.46	+.90	2	1	0
4. Abstract	+.18	+.10	+.24	+.85	1	4	1
5. Men	+.23	+.16	+.51	+.87	1	0	0
6. Groups II	+.27	+.16	+.53	+.78	3	0	0
7. Religious II	+.31	+.23	+.40	+.92	0	4	0
8. Distorted or Fantastic	+.28	+.24	+.35	+.87	0	1	1
9. Women	+.29	+.23	+.41	+.89	1	0	0
10. Children	+.43	+.38	+.52	+.94	1	0	0
11. Animals	+.32	+.18	+.58	+.87	1	0	0
12. Still Lifes	+.35	+.19	+.52	+.89	0	3	0

relation to this of each individual's ordering. A simple way of obtaining the group ordering would be to calculate for each picture the mean preference rating given it by the 22 Ss in a group. A rather better way and almost as easy in these days of electronic calculators, is to use factor analysis. It enables us to determine for each picture a mean preference rating in which each S's rating is weighted according to the degree to which the S agrees with the group ordering of the particular set of pictures. The saturation of each S on the first factor, in a factor analysis of the inter-person correlation matrix, provides a measure of the extent to which that S agrees or disagrees with such a group ordering. If his tendency is to agree, the saturation will be positive; if to disagree, the saturation will be negative. Table 1 shows in columns E and F how many of the Ss in each group, for each set of pictures, show disagreement rather than agreement. It is clear that some tendency toward agreement is very widespread. There are 24 opportunities to test the question, each represented by an entry in column E or F. In 10 instances, every one of the 22 Ss in the group has a positive saturation and in 9 more instances only one S has a negative saturation.[2] The tendency to agree with others in personal preferences, for most of the sets of pictures used, characterizes nearly all of the Ss.

A third way of testing agreement is to ask whether the group ordering of a set of pictures by college students in one part of the country resembles the group ordering by college students from other parts of the country -- though at a similar kind of institution. The answer, shown in column D of Table 1, is that the resemblance is very close. In all but one set of pictures the correlation between the two orderings was between +.84 and +.94: the exception fell to +.78.

Consistency of Individual Differences in Agreement with Group Preferences

For each set of pictures, the saturation of each S on the first factor provides a measure of the extent to which he agrees with the group preference ratings. The meaning of this measure depends first of all on whether an individual S tends to have a consistent position on this measure for different sets of pictures.

A first indication of degree of consistency is provided by the fact that two kinds of pictures are each represented by two sets. Each set yields for each S, of course, a separate first-factor saturation. The measures so derived from the two sets of paintings of groups of people are correlated with each other to the extent indicated by a coefficient of +.87 for the Stanford data and +.53 for the Yale data. The measures derived from the two sets of religious pictures yield a correlation coefficient of +.73 for the Stanford data and +.62 for the Yale data.

---

[2] The negative saturations that do occur, moreover, tend to be small. The 25 negative saturations have a median value of -.20, whereas the 503 positive saturations have a median value of +.54.

Of greater interest is the stability of this measure when the kind of picture is varied. For the 10 kinds of pictures used (omitting the second set of religious and of group paintings), all the intercorrelations have been determined for the Stanford data, and for the Yale data all 12 sets have been used. The median correlation is +.49 for the Stanford data and +.48 for the Yale data. The correlations are very low for the abstract pictures in relation to other sets, and one such correlation (out of 20) is even negative. Otherwise, the indication throughout is of consistency. The extent to which an S agrees with the consensus is a reliable individual characteristic.

Group Preferences and the External Criterion of Esthetic Value

For each set of pictures, an external criterion of esthetic value was provided by expert judgment, as described above. The experts, asked to judge esthetic merit, agreed with each other decidedly better than did the students asked to express personal preferences. Their median correlation coefficients are presented in column C of Table 1 for comparison with those from the preference data. For the merit judgments, negative correlations between people were very rare (4 per cent of the correlations as against 16 per cent for the preference data), and they were concentrated in the abstract and distorted-fantastic sets. Whereas every set of pictures had a considerable number of negative correlations among Ss' preferences, 6 of the 12 sets had no negative correlations whatever for judgments of merit. In a factor analysis of these interjudge correlations, negative saturations on the first factor (indicative of disagreement) were extremely rare, as indicated in column G of Table 1. Only one judge had any negative saturations, and only for those sets of pictures -- abstract and distorted-fantastic--for which agreement was generally poorest. These two negative saturations were small (-.23 and -.17), whereas the 15+ positive saturations had a median value of +.71.

The measure of esthetic value for each picture emerged from a factor analysis of these judgments of merit, and it represents the average rating by the 13 judges, with each judge's rating being weighted according to how good a judge he was on the evidence of his agreement with the other judges in evaluating the given set of pictures.

This ordering of the 60 pictures for esthetic value could then be compared with the preference ordering by each of the two groups of Ss. For each set, the resemblance of the orderings may be expressed by a correlation coefficient. The results are presented in Table 2. The tendency is toward a positive relationship, but a very

Table 2

Correlation Between Ordering of Pictures by Group Preference and by Aesthetic Value

Set of pictures	Stanford	Yale
1. Groups 1	+.02	+.11
2. Religious I	+.25	+.31
3. Landscapes	+.14	+.28
4. Abstract	+.01	-.10
5. Men	+.34	+.51
6. Groups II	+.04	+.29
7. Religious II	-.06	-.16
8. Distorted or Fantastic	-.04	-.04
9. Women	+.26	+.33
10. Children	+.23	+.39
11. Animals	+.09	+.31
12. Still Lifes	+.02	+.02

small one; the median coefficient is +.125. The ordering of a set of pictures by the preferences of a group of unselected students, then, has generally very little to do with their esthetic value.

Consistency of Individual Differences in Agreement with the External Criterion of Esthetic Value

When, for a given set of pictures, each individual's order of preference was considered in relation to the order of esthetic value derived from expert judgments, the relationship was of course found to vary from one S to another. Consider for example the 44 different Ss' ratings of the pictures of Set 1. At one extreme, one S's preferences showed a correlation of +.69 with the order of esthetic value, and at the other extreme another S's showed a correlation of -.29.

Thus for each S 12 separate measures were available of the extent to which his preferences among paintings were related to the criterion of esthetic value -- one measure from each set of 60 pictures. The intercorrelations among all these 12 measures have not been determined. Instead the 12 were divided into two sets of 6, for each of which an average has been taken for each S. The correlation was then determined between these two halves of the total measure available. For the Stanford group the correlation was +.96, and for the Yale group it was +.90. Clearly, given the kinds of pictures used here, the degree to which preferences are related to esthetic value is a very stable characteristic of the individual. The comparable split-half correlations for the degree to which individual preferences are related to group preferences were +.87 and +.80 for the Stanford and Yale groups, respectively. As far as this evidence goes, degree of agreement with an esthetic standard is an even more consistent characteristic than degree of agreement with group preferences.

Relation Between the Two Measures of Individual Preference

It has been reported above that the ordering of pictures by group preference and their ordering by judgments of esthetic value tend to be positively related, but only to a very small degree. When individuals are measured for the extent to which their preferences resemble the one kind of standard or the other, a very different relationship appears for the two bodies of data. In the Yale data, persons whose preferences most resemble the group ratings have preferences which least resemble the criterion of esthetic value, and vice versa, to the extent indicated by a correlation coefficient of -.79. For the Stanford data, there simply appears to be no relation between the two measures; the coefficient of correlation is -.03.

These results are entirely contrary to the assumptions and interpretations made in much of the earlier research referred to in the introduction. In a group of persons selected without reference to interest in art -- at least in such groups selected in American universities -- tendency to agree with the consensus in picture preferences seems to have a variable significance. But it certainly is not a dependable positive measure of esthetic sensitivity.

Relations with Other Features of Personality and Background

Several other measures of reactions to art, of experience with art, and of other personal characteristics of the Ss were obtained or were already available. For 20 measures available for both groups of Ss, the correlation with both the preference measures was obtained, separately for each group. The results are given in Table 3 for all variables for which at least one of the correlations is significant at the 5 per cent level. The general pattern is what might be expected from results already cited: degree of agreement with esthetic value has approximately the same correlations for the two groups of Ss. Degree of agreement with group preference tends to have correlations opposite to these; this opposition is generally more marked for the Yale Ss than for the Stanford Ss.

About the significance of agreement with esthetic value, it seems safe to draw some tentative conclusions from these results. First, such agreement is part of a general complex of knowledge about art and skill in evaluating it.

The highest correlation of all is with goodness of esthetic judgment in the Yale Ss, as measured by those items on the Bulley Test (Bulley, 1951) which had proved discriminating in a prior trial of the entire test with Yale undergraduates; the corresponding correlation for the Stanford Ss is also very substantial. The Bulley Test consists of pairs of pictures similar in subject matter or type, selected to be of differing esthetic merit; the S is asked explicitly to make an esthetic judgment, indicating which member of each pair he believes to be the better work of art. With respect to individual differences, therefore, the Bulley Test provides an external criterion of esthetic sensitivity, or goodness of esthetic judgment. The results obtained with it provide independent evidence that degree of agreement between one's preferences and esthetic value is an index of esthetic sensitivity. To anticipate a bit, the results also show that for the Ss of the present study, degree of agreement with group preferences does not correspond to an external criterion of esthetic sensitivity.

In both groups, agreement of preference with esthetic value is correlated with information about art, as rated from the interview. It is correlated also with a tendency for the S, when asked, about the bases for his preferences in this research, to mention criteria of esthetic value. For one group, at least, agreement of preference with esthetic value is also correlated with a measure of stability of

Table 3

Agreement of Individual Preferences with Group Preference and with Aesthetic Value: Relation of These Measures to Other Variables

(Product-moment correlations)

Variable	Degree of Individual Agreement with			
	Group Preference		Aesthetic Value	
	Stanford	Yale	Stanford	Yale
Variables dealing with art				
1. Interview rating: extent to which bases for preference referred to aesthetic criteria	-.22	-.41	+.66**	+.48*
2. Interview rating: information about art	-.05	-.47*	+.76**	+.70**
3. Stability of aesthetic preference	+.02	+.07	+.51*	+.1
4. Goodness of aesthetic judgment (abbreviated Bulley Test)	-.06	-.55**	+.56**	+.85**
5. Barron-Welsh Art Scale	-.19	-.37	+.45*	+.32
Sentence-Preference Test: liking for paradox, irony, and ambiguity in everyday language	-.15	+.38	-.35	-.57**
Sentence-Preference Test: liking for other poetic characteristics in everyday language	+.42	+.23	.00	-.53*
Scholastic Aptitude Test, Verbal	+.18	-.21	+.53	+.45*
Intuition as contrasted with Sensation (Myers-Briggs Type Indicator)	-.44*	-.08	+.28	+.11
Perception as contrasted with Judgment (Myers-Briggs Type Indicator)	-.52*	-.26	+.38	+.22
Viscerotonio (questionnaire)	+.33	+.42	-.27	-.47*
Tolerance for ambiguity and ambivalence (questionnaire)	-.03	-.62**	+.26	+.64**

* $p < .05$.

** $p < .01$.

esthetic preference. This measure was based on the agreement between the ordering of a sample of 6 pictures of each of the kinds used when presented in a mixed set at the end of the preference series, and the original ordering when the samples of 6 pictures had been included in their own sets. For both groups, agreement of preference with esthetic value is correlated somewhat with a tendency to prefer, among abstract designs, kinds of designs selectively preferred by artists and by original people generally (Barron-Welsh Art Scale: see Barron & Welsh, 1952; Barron, 1958; Welsh, 1959).

The Sentence-Preference Test, listed after the art variables in Table 3, is an unpublished test devised by the author and his students whose meaning is too uncertain to justify a confident interpretation. It is possible that a liking for poetic elements in everyday language represents poor esthetic judgment in world art, and that there is a correlation between esthetic judgment there and in visual art. But this hypothesis would need more direct testing.

The substantial correlation with the verbal score on the Scholastic Aptitude Test probably has a complex meaning. At least we may conclude that agreement of preference with esthetic value is positively related to general intellectual adequacy in the verbal sphere; it gives no sign of being an adoption of tradition by the less able for the sake of security.

There is some evidence, stronger for the Stanford group than for the Yale group, but in neither case significant, that agreement of preferences with esthetic value is correlated with intuition rather than sensation, and with perception rather than judgment, among the Jungian variables measured by the Myers-Briggs Type Indicator (a questionnaire at present distributed for research purposes by the Educational Testing Service, Princeton, N.J.).

The negative correlation with viscerotonia and the positive correlation with tolerance for ambiguity and ambivalence, both as measured by a number of questionnaire items, are sizable only for the Yale group, but present also in the Stanford group. Viscerotonia, a concept originating with Sheldon (1942), refers to a cluster of temperamental characteristics which seem to center around love of comfort, relaxation, and friendly social relationships. It was measured here by questionnaire items based on those found by Child (1956) to be correlated with physique in the way asserted by Sheldon. Tolerance for ambiguity and ambivalence was measured by questionnaire items based on apparently relevant items in the F scale of Adorno, Frenkel-Brunswik, Levinson, & Sanford (1950); to escape the influence of a response set half the items were written so that tolerance would be expressed by accepting rather than rejecting the statement. The correlations with viscerotonia and with tolerance for ambiguity are consistent with the positive relation to verbal aptitude and to the Barron-Welsh Art Scale (which has been used by others as an index of originality) in suggesting that a positive relation between preferences and esthetic value tends to arise out of ability, creativity, and direct interaction with a complex object, more than out of insecurity, need to accept the dictates of authority, and habitual conformity.

As for the other measure, degree of agreement with group preference, the results differ more for the two groups of Ss, and the conclusions must be even more tentative and limited. (a) Its personal significance seems likely to vary a good deal from one group to another. (b) For the two groups drawn on in this study, its relationship to other variables never seems to resemble that of the degree of agreement with esthetic value. Whenever one of these two measures is appreciably related to a variable, the other measure appears either to be related to that variable in the opposite direction, or to show no appreciable relation to it.[3]

DISCUSSION

In previous studies referred to in the introduction, ordering of stimuli by group preference has been assumed to be an ordering by esthetic value, and the closeness of an individual's agreement with this group preference order has been assumed to be a positive measure of esthetic sensitivity. In the present research, these assumptions have been checked against an external criterion of the esthetic value of the stimuli. The first assumption has been confirmed only in very slight degree and not for all the kinds of pictures used. The second assumption has been found, for the present samples, to be completely invalid. If esthetic sensitivity is expressed in a tendency to prefer the esthetically good, then agreement with group preference is in one group negatively related to esthetic sensitivity and in the other group unrelated.

For the second assumption, that individual agreement with group preference is a positive measure of esthetic sensitivity, this research also provides a second check, entirely independent of the first, through inclusion of the Bulley Test of esthetic judgment. The results of this second check of the assumption are in perfect agreement with those just summarized. For one group of Ss, as is shown in Table 3, agreement with group preferences shows almost no relation to this independent measure of esthetic sensitivity ($r = -.06$); for the other group, the relation is decidedly negative ($r = -.55$).

---

[3] With the following 8 variables, available for both groups of Ss, no significant correlations were obtained with either preference measure: (1) extroversion vs. introversion and (2) feeling vs. thinking on the Myers-Briggs Type Inventory; (3) extroversion vs. introversion, (4) somatotonia, and (5) anxiety, on a questionnaire assembled for this research; (6) time required to express the painting preferences; (7) mathematical score on the SAT; (8) human movement score on the Barron Inkblot Test (Barron, 1955).

How did these assumptions which are apparently false come to be so widely accepted? Initially, perhaps, because they were first made in connection with preferences among simple stimuli such as odorants and color patches, where an external criterion of esthetic value does not readily offer itself, against which to test the assumptions. (In the present research, a small-scale effort was made to get judgments of esthetic value on simple visual stimuli. The agreement was so much less than on the paintings that a large-scale attack would be necessary to get a stable measure, if indeed one is possible.) Then in the small amount of previous research which used stimuli of more likely esthetic import, the assumption was not adequately tested. H. J. Eysenck (1940), in reporting the most relevant example of such research, does not give any facts which permit one to judge whether the persons agreeing most closely with the group preferences also agreed with an external standard of esthetic value. They may have, of course. If so, the present research suggests a possible reason for the discrepancy. Some individuals' preferences are closely related to esthetic value. If such individuals comprised a large enough portion of the Ss in a study, their preferences would have a controlling influence on the average preferences of the group. A group composed either largely or entirely of such persons should indeed yield a positive correlation between agreement with group preferences and agreement with esthetic norms. Eysenck's description of his sample leaves this possibility open though it by no means establishes it.

Eysenck (1940) used, within a set, pictures which were selected to meet the following criteria: "No tradition or teaching should point to one of them as superior to the others; so far as execution is concerned they should be roughly of the same degree of excellence; and they should all be equally unknown to the subjects." These criteria led him to such sets as pictures of medieval clocks, colored photographs of various flowers, photographs of modern steamships, pictures of modern statues, and reproductions of drawings by a single artist. In general, the pictures within a set in his research probably varied in esthetic value, and in other sources of appeal, much less than was true for those used in the present research. The same criteria used by Eysenck might be said to be met on the whole, by the sets of colors or smells used in other previous research. These criteria are not met in the present research; on the contrary, pictures were used which differed radically from each other in all three respects cited in the quotation from Eysenck. It is conceivable that the greater variety of pictures in a set, with the greater opportunity for a number of factors to influence preferences, might actually reverse the direction of relationships found, but this seems unlikely.

The lack of external criteria in the previous research, however, leaves open the possibility that these apparently erroneous assumptions may have been erroneous even in the specific context of that research. They may have been a valuable first guess at the meaning of what is indeed a very striking phenomenon--the very considerable consistency in the degree to which a given person's preferences agree with group tendencies which appear to be extremely diverse in their nature. To the understanding of this phenomenon I fear this report makes principally a negative contribution.

But if personal preferences do not reveal esthetic sensitivity through agreement with group preferences, they may do so through agreement with an external standard of esthetic value provided by the judgment of experts.[4] In the present research, individual differences in the agreement of preferences with an external criterion of value are at least as stable as are differences in agreement with the group.

What is the meaning of this variation among people in the extent to which their preferences are correlated with a standard of esthetic value? Until much cross-cultural and developmental research not yet completed has been done, we cannot confidently judge its meaning. Meanwhile, some will assume it to signify simply variation in the extent to which people are familiar with and accept a particular tradition of esthetic evaluation. That it has much to do with such familiarity and acceptance can hardly be doubted. But I for one would be inclined to guess that various traditions of esthetic evaluation have a good deal in common, attributable in part to objective facts about the esthetic potentialities of various works of art when observed with understanding, and that a careful observer is not completely dependent upon familiarity with a tradition to discover some of those facts for himself. For one who entertains this hypothesis, the results of this research suggest that agreement of personal preferences with a standard of esthetic value may be related to certain variables of intellectual approach and temperament which affect the person's openness to the esthetic potentialities of his environment.

SUMMARY

Previous authors have suggested that the average preference rating of an esthetically relevant stimulus be taken as a measure of its esthetic value, and that the degree to which a person agrees with the average preference ratings of various stimuli be taken as a measure of his esthetic sensitivity. This paper reports an investigation of the empirical consequences of these suggestions, using as stimuli reproductions of paintings, for which an external

---

[4] I am not arguing that personal preferences provide the best technique for measuring esthetic sensitivity. If one set out to measure esthetic sensitivity, he would ordinarily not ask people to express personal preferences; he would generally do better to ask them to make esthetic judgments, as in the Bulley Test. I only intend to argue that if one is dealing with personal preferences (because they are the data available), they may yield a usable measure of esthetic sensitivity if they are evaluated by reference to an external criterion of esthetic value, but not if they are evaluated by reference to average preference.

criterion of esthetic value is available in judgments made by art experts. For two groups of college men unselected for competence in art, average preference ratings of paintings bore little relation to the esthetic value of the paintings. Individual degree of agreement with average preferences showed in one group a negative relation, and in the other group no relation, to measures of esthetic sensitivity, whether the latter were based on agreement of preferences with the criterion of esthetic value or on an independent measure provided by the Bulley Test of esthetic judgment. Other tests and interviews were then drawn on to investigate personality correlates of esthetic sensitivity.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D.J. & Sanford, R. N. The authoritarian personality. New York: Harper, 1950.

Barron, F. Threshold for the perception of human movement in inkblots. J. consult. Psychol., 1955, 19, 33-38.

Barron, F. The psychology of imagination. Sci. Amer., 1958, 199, 150-166.

Barron, F., & Welsh, G.S. Artistic perception as a possible factor in personality style. J. Psychol., 1952, 33, 199-203.

Beebe-Center, J.G. The psychology of pleasantness and unplesantness. New York: Van Nostrand, 1932.

Bulley, Margaret H. Art and everyman, Vol. I. London: Batsford, 1951.

Burt, C. The psychology of art. In C. Burt et al., How the mind works. London: Allen and Unwin, 1933. Pp. 267-310.

Child, I. L. The relation of somatotype to self-ratings on Sheldon's temperamental traits. J. Pers., 1950, 18, 440-453.

Eysenck, H.J. The general factor in esthetic judgments. Brit. J. Psychol., 1940, 31, 94-102.

Eysenck, H. J. A critical and experimental study of colour preferences. Amer. J. Psychol., 1941, 54, 385-394.

Eysenck, H. J. Sense and Nonsense in Psychology. Hammondsworth and Baltimore. Penguin Books, 1957.

Eysenck, Margaret D. An experimental and statistical study of olfactory preferences. J. exper. Psychol., 1944, 34, 246-252.

Granger, G.W. An experimental study of colour preferences. J. gen. Psychol., 1955, 52, 3-20.

Pickford, R. W. Factorial studies of esthetic judgments. In A. A. Roback (Ed.), Present-day Psychology. New York: Philosophical Library, 1956. Pp. 915-929.

Sheldon, W. H. The varieties of temperament. New York: Harper, 1942.

Valentine, C. W. The experimental psychology of beauty. London: Methuen, 1962.

Welsh, G. S. Preliminary manual, Welsh Figure Preference Test. Palo Alto: Consulting Psychologists Press, 1959.

Chapter II

OBSERVATIONS ON THE MEANING OF SOME MEASURES OF ESTHETIC SENSITIVITY

I. INTRODUCTION

Esthetic judgment, esthetic preference, and esthetic sensitivity are concepts having to do with evaluation, by reference to some external standard of esthetic value, of a person's responses to stimuli to which that standard may be applied. Distinctions among these three concepts are not generally made, and the distinctions are useful. They will be made here with reference to the procedures of psychological testing, but it is obvious that equivalent distinctions can be made for the phenomena of everyday behavior.

Esthetic judgment, as a measured variable, is the extent to which, when a person judges the esthetic value of stimuli, his judgments correspond to the external standard of value which is being employed.

Esthetic preference, as a measured variable, is the extent to which, when a person expresses (by word or action) his relative liking or disliking of various stimuli, this relative preference for various stimuli corresponds to their esthetic value as defined by the external standard.

Esthetic sensitivity is a more inclusive term. It refers to the extent to which a person gives evidence of responding to relevant stimuli in some consistent and appropriate relation to the external standard. Esthetic judgment and esthetic preferences may be viewed as special cases of esthetic sensitivity. If a person gets a high score on a test either of esthetic judgment or of esthetic preference, he may be thought of as displaying a high degree of esthetic sensitivity. If he gets a low score on a test of esthetic judgment, he may reasonably be said to show low esthetic sensitivity (at least with respect to the particular standard of esthetic value), provided that he was a cooperative subject. If a person gets a low score on a test of esthetic preference, the case is not so clear; there is some probability that he

---

1. It is also obvious, though, that the concepts defined here do not correspond exactly to all the similar concepts used in other contexts and potentially capable of psychological study. For example, good esthetic judgment may be ascribed to a person more because of the reasons he gives in evaluating a work of art, than because of his agreement with an external standard.

is low in esthetic sensitivity, but there are a variety of other possible reasons for the low score besides the one suggested by this label. He may be quite sensitive esthetically (as measured, say, by a test of esthetic judgment) and yet base his preferences primarily on non-esthetic considerations (e.g., monetary value, relevance to his emotional state, etc.).[2]

These concepts (sometimes under other labels, such as "goodness of taste") have in the past often been used in a potentially quite culture-bound context, in either of two ways:

(1) Traditions of evaluation of art in our society are assumed to have an absolute validity. These traditions, as embodied in the evaluative behavior of one or more expert critics, are used in selecting or constructing stimulus materials for a test. The test calls for the examinee to make judgments about esthetic value, or to express preferences, and the extent to which he agrees with the judgments or preferences of experts is taken as a measure of his esthetic sensitivity. Examples are to be found in the art aptitude tests constructed by Meier (1940) and by Graves (1946), and in the materials for testing esthetic judgment assembled by Bulley (1951, and her earlier writings referred to therein).

(2) When people are asked to rank a number of stimuli according to personal preference, the consensus[3] might be thought of as the order of preference that tends to characterize everyone to some extent, but is in an individual's ranking partially obscured by idiosyncratic tendencies. The extent to which any one person's preferences agree with the consensus may then be taken as an indication of the extent to which he is sensitive to the presumably universal esthetic appeal of stimuli--as a measure, in short, of esthetic sensitivity. Where this argument has been presented, it seems to have been accompanied by a naive, unstated assumption that a sample of subjects obtained in a Western metropolis will

---

[2] There are possible ambiguities in a high score, too. Perhaps a high scorer is really insensitive to esthetic values, but has learned to recognize the works highly valued by critics and to express a preference for them, or to say they are better. If this ambiguity arises even for a well-constructed test of esthetic preference or esthetic judgment, then it might be useful to define and measure esthetic sensitivity in some very different way.

[3] Consensus is used in this paper to mean the common tendencies, the agreement found, in the reactions of a number of people to a set of stimuli, with no implication that they have arrived at agreement through interaction with each other. In studies such as are mentioned here, the reactions of subjects are obtained independently, so that the sources of any consensus do not lie in social interaction among subjects during the testing (nor before, generally, except in the sense of their separate participation in a single community and society).

provide an adequate basis for defining the pan-human consensus. This use of the concept is to be found in the writings of Beebe-Center (1932), Eysenck (1940a, 1957), and Granger (1955a, 1955b).

These concepts need not, however, be set in a culture-bound context. There are two ways in which they may be freed.

(1) Whether the standard against which responses are evaluated is provided by expert judgment or by a consensus of ordinary people, it is possible that this standard might turn out to have some trans-cultural validity. In the case of the popular consensual standard, the little research that has been done on visual designs or pictures (McElroy, 1952; Lawlor, 1955) casts great doubt on this possibility. No systematic research seems yet to have been done on the trans-cultural validity of expert judgments. An important degree of trans-cultural validity seems to me more likely for expert judgment than for popular preference, in the case of works of art. (In the case of certain simple stimuli--odorants, at least--substantial trans-cultural validity seems highly likely for popular preference, too, though whether reactions to such stimuli have anything to do with reactions to art is another matter.) Agreement with a standard derived from within our own society could thus provide tentative evidence of agreement with a possible universal tendency (assumed here to be testable, and ultimately to be tested, not assumed to be given).

(2) The judgment of experts or the consensus of ordinary people may be thought of as culturally determined, as probably having no trans-cultural validity, and yet as providing a meaningful standard for people living within the same culture. Agreement with such a standard, it could be argued, shows that one is sensitive to the esthetic value that stimuli potentially have for people participating in that culture. Esthetic sensitivity might then have some trans-cultural meaning, even though the standards for evaluating it would vary from one society to another. (In the same way, we might suppose intensity of theological belief to be a trans-cultural variable, though in one society the intense believer is a monotheist and in another a polytheist.)

Much that has been written about esthetic sensitivity sounds antiquated to the psychological reader of today, because it is expressed in the naively culture-bound phraseology that came naturally to many Westerners, even in the social sciences, until recent years. But the reader of today should not be misled by this context into thinking that the problems dealt with in those writings are in any way settled by our decreasing ethnocentrism. On the contrary, they are more alive than ever because of the increasing possibility of bringing them to test.

In the course of beginning a program of research on esthetic judgment, I did some exploratory data-gathering with some of the measures of esthetic sensitivity that had been originally devised in a more culturally naive setting. The data obtained are helpful in answering some questions, and in raising others, about some of these measures.

II. TWO TESTS OF ESTHETIC SENSITIVITY

The first data I collected involved the use of two tests of esthetic sensitivity. One is the Design Judgment Test devised by Maitland Graves (1946), which in the terminology used here should be considered as a test of esthetic preference (since the subject is asked to express a preference, not to make a judgment). The subject must choose between the drawings in a pair (in some instances, a trio), where the two drawings--simple abstract designs rather than works of art--are of similar type but differ in a way that can be rather easily conceptualized. The author of the test provides, in a separate book (1941), the conceptualization that guided his construction of the test. The manual (1948) published with the test reports high reliability, shows that mean scores of art students and non-art students differ greatly, and indicates that items were selected from a larger pool on the basis of expert judgment, group discriminatory value, and consistency with total score. Through the courtesy of the copyright owners I was able to make slides and project the stimuli on a screen, obtaining standard exposure time in presentation to a large class.[4]

The second test is the one described by the English esthetician, Margaret Bulley, in her book Art and Everyman (1951), and available in filmstrip from the publishers of that book. It presents pairs of art objects and asks the subject to judge which work in the pair is the better work of art. Data are given on group differences, both on individual items and on total score, which confirm the general validity of the test. Such information seems not to have been used to improve the test, however. The author evidently intended the test principally for demonstrational purposes rather than for research or practical purposes, and was content to rely mostly on her own judgment in justifying the choice of items. Despite the mystical phraseology in which the author couches her statements, this test seems to have more obvious face validity than analytic tests such as Graves. It asks for esthetic judgments about actual works of art. Such a test has better initial claim to measure esthetic sensitivity than a test whose stimuli are not works of art, and whose instructions do not direct attention to esthetic values.

These two tests were given as a class demonstration in an undergraduate course in psychology at Yale University. 105 students (all men) were present for both tests, and some additional students for only one. It was apparent that some of the items in the Bulley Test did not yield results consistent with the other items. A simple item analysis in relation to original total score was performed, and the 19 best items (out of 32 originally used) were the basis of all subsequent results to be reported here.

---

[4] I am indebted to The Psychological Corporation for permission to make these slides for use in research.

A first result of interest is that these two tests have very little relation to each other; the correlation between them is only +.12. This is an extreme instance of the general finding that various standardized tests of esthetic sensitivity are not very highly related to each other. For intercorrelations from +.21 to +.51 among various such tests, see Anderson (1951), Barrett (1949), Bolton (1955), Carroll & Eurich (1932), Crannell (1953), and Dewar (1938).

If two tests purporting to measure the same variable are so little related to each other, then in judging what each measures we are especially dependent upon knowledge of what other variables it is related to. For varying numbers of persons, information was available on 17 other measures of potential relevance. Correlations with these are given in Table 1. The first 12 variables in the table pertain to academic aptitudes and achievement. The results with most of these provide no clear evidence of a relation with either the Bulley or the Graves test. The Bulley test does have one decidedly significant relationship, a correlation of +.36 with a Spatial Relations Test. The Graves test shows a suggestion of positive relationship to a series of achievement tests in the sciences and humanities (variables 9-12), though only one of the correlations is barely significant. The last 5 variables are based on questionnaires about background in and reactions to art. The Bulley test shows significant positive correlations with the extent to which a person reports having had experience with visual art (variable 17) and with a self-rating of adequacy of esthetic judgment in visual art (variable 13). The Graves test shows no significant relation with these or with the remaining variables pertaining to art. As far as the evidence goes, then, it suggests that the Bulley test is somewhat related to other measures of esthetic sensitivity or of background likely to be correlated with esthetic sensitivity, and it fails to offer any corresponding suggestion for the Graves test. But the evidence is not very conclusive; the questionnaires were taken after both tests but in closer proximity to the Bulley than to the Graves test, and answers might conceivably have been influenced by a private feeling of adequacy or inadequacy in responding to the Bulley test (though this seems unlikely for the relatively objective information about experience with visual art). Further evidence tending to support these suggestions of difference in the meaning of the two tests will, however, be offered in the following section of this paper.

A final point of interest emerges from consideration of the internal consistency of each of these tests. For this purpose, attention will be directed only at those items in the Graves test (82 out of the total of 90 items) which, like all of the items in the Bulley test, require a choice between just two alternatives. Suppose that a test composed of such items is split into halves,

Table 1

CORRELATION OF BULLEY AND GRAVES TESTS WITH OTHER VARIABLES
(number of subjects given in parentheses)

	CORRELATION WITH			
VARIABLE	BULLEY SCORE		GRAVES SCORE	
1. Scholastic Aptitude Test, Verbal	+.05	(122)	+.18	(110)
2. Scholastic Aptitude Test, Mathematical	-.02	(122)	-.00	(111)
3. Verbal Reasoning Test, Yale Aptitude Battery[1]	+.07	(105)	+.05	(96)
4. Quantitative Reasoning Test, Yale Aptitude Battery	+.04	(105)	+.00	(96)
5. Spatial Relations Test, Yale Aptitude Battery	+.36**	(52)	-.04	(44)
6. Mechanical Ingenuity Test, Yale Aptitude Battery	+.10	(40)	-.00	(36)
7. Grade Average, Freshman Year	-.06	(121)	+.16	(110)
8. English Achievement Test	+.12	(97)	+.16	(88)
9. Test of Developed Ability: Science Knowledge[2]	-.12	(60)	+.14	(54)
10. Test of Developed Ability: Science Abilities	-.23	(60)	+.22	(54)
11. Test of Developed Ability: Humanities References	+.12	(60)	+.24	(54)
12. Test of Developed Ability: Humanities Abilities	+.07	(60)	+.30*	(54)
13. Self-rating of goodness of esthetic judgment in visual arts	+.39**	(129)	+.16	(105)
14. Self-rating of goodness of esthetic judgment in other arts	+.19*	(129)	+.15	(105)
15. Belief in objectivity of esthetic values in visual arts	+.04	(128)	-.06	(104)
16. Belief in objectivity of esthetic values in other arts	+.12	(128)	+.10	(104)
17. Experience with visual arts (rating based on answers to several questions)	+.42**	(107)	+.12	(96)

---

* $p<.05$

** $p<.01$

[1] For information about the Yale Aptitude Battery, see Crawford & Burnham (1946).

[2] For information about the Tests of Developed Ability, see Coffman (1957) and Dyer (1957).

and a score for each person is determined separately for each half. If the test is a reliable measure of esthetic sensitivity, what should be found? At the upper end, the answer is clear: some individuals should show a consistent tendency to make the "right" choices on each half. But at the lower end, the expectation is not so clear. Perhaps it is most reasonable to expect that the low-scoring individuals should show no consistent tendency at all, that on the two halves of the test their scores should be scattered at random around the 50% level that might in the long run result from failure to discriminate in any consistent way. (To state a parallel case: In an intelligence test consisting of two-choice items, a testee required to respond to all items but lacking the ability to cope with them would be expected on the average to give 50% correct answers. If stupid testees gave consistently less than 50% correct answers, the test would be considered defective, for no one should be able to discriminate consistently between the right and the wrong answer except on the basis of psychological processes which would at the same time tell him which is right and which is wrong.)

Figure 1 shows that for the Graves test this expectation is not fulfilled. There clearly are subjects who are consistent in tending to prefer the design which in the scoring is regarded as the poorer choice, and deviation from the chance level is nearly as great in this direction as it is in the opposite direction, though much less frequent.

For the Bulley test comparable findings were obtained. They will not be presented, since the lack of prior selection of items would leave these findings rather ambiguous. If the entire test were used, the results would include items which show no consistency with the rest. If only the 19 selected items were used, the consistency below as well as above the chance value would be in part a result of sampling error.

Persons who score consistently low on one of these tests are given no special consideration by the authors of the tests. A low score is regarded as indicating a lower degree of esthetic sensitivity than is indicated by a medium score. Against this position, it may be objected that the low scorers are in a sense showing a higher degree of esthetic sensitivity than the medium scorers. They are consistent in discriminating among stimuli on the same basis as do the high scorers. They differ from the high scorers in the evaluation which follows upon the discrimination, but they appear at least superficially to be sensitive to the same aspects of the stimuli as are the high scorers.

From a practical point of view, this characteristic of the Graves and Bulley tests would be extremely limiting. Certain testees could, just by instructing themselves to respond with the opposite of their spontaneous choices, convert themselves from low scorers to high scorers.

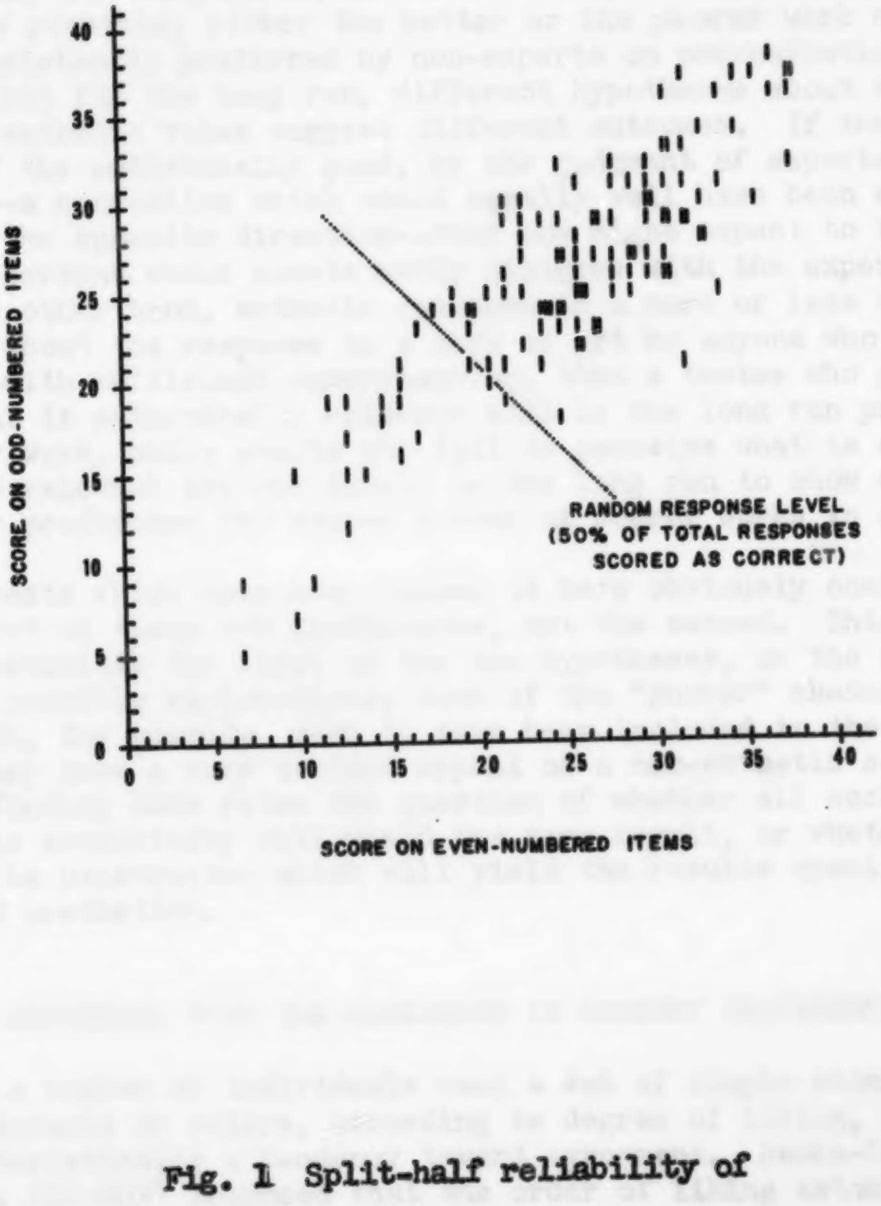

Fig. 1 Split-half reliability of two-choice items in Graves Design Judgment Test.

The greatest interest of these findings on internal consistency, though, lies in the questions they raise about the nature of esthetically good and bad art. The practical limitation just mentioned may result from special characteristics of the materials chosen for these particular tests, and have no general significance. Going beyond the materials of these particular tests, one may ask, Can works of art in general be dependably sorted, into those which experts consider good and bad, by a person who does not agree with this evaluation? Or are the differences between good and bad such that anyone who has come to discriminate these differences must inevitably tend in the long run to agree with the experts on the evaluation which follows upon this discrimination? It is obvious that for any one comparison, or set of similar comparisons, any outcome is possible; either the better or the poorer work of art may be consistently preferred by non-experts on non-esthetic grounds. But for the long run, different hypotheses about the nature of esthetic value suggest different outcomes. If the definition of the esthetically good, by the judgment of experts, is arbitrary--a convention which could equally well have been established in the opposite direction--then one might expect to find that some persons would consistently disagree with the experts. If, on the other hand, esthetic goodness is a more or less objective fact about the response to a work of art by anyone who observes it with sufficient understanding, then a testee who perceives what is esthetically relevant will in the long run prefer the better work, while people who fail to perceive what is esthetically relevant are not likely in the long run to show any consistent preference for either better or poorer works in general.

The tests which have been looked at here obviously conform to the first of these two predictions, not the second. This fact does not establish the first of the two hypotheses, as the fact has other possible explanations; some of the "poorer" choices in the Bulley test, for example, seem to have been included in the test because they have a very decided appeal of a non-esthetic sort. But this finding does raise the question of whether all such tests of esthetic sensitivity will yield the same result, or whether tests can be constructed which will yield the results specified in the second prediction.

III. AGREEMENT WITH THE CONSENSUS IN SENSORY PREFERENCES

When a number of individuals rank a set of simple stimuli, such as odorants or colors, according to degree of liking, there is characteristically a tendency toward agreement. Beebe-Center (1932, pp. 202-203) proposed that the order of liking established by consensus might indicate an order of general esthetic value, and that the extent to which any individual's order of preference agreed with the consensus might be a measure of his esthetic sensitivity. Beebe-Center did not elaborate upon this suggestion, and it is not clear whether he supposed esthetic sensitivity, so defined, to be a general variable which would be related to esthetic sensitivity as ordinarily thought of in connection with reactions to works of art. But exactly this supposition was made in later elaboration of this idea by several English psychologists, well represented by Eysenck (1957).

There has, moreover, been some evidence supplied to support this idea. The most persuasive evidence is probably that offered by Granger. In one study (1955a) he determined extent of individual agreement with the consensus in preference rankings of 60 different sets of Munsell color chips (each set containing an average of 7 chips, generally differing only on one color dimension--hue, saturation, or brightness). This measure, available on 50 subjects, was then correlated with scores on the Graves Design Judgment Test, and a coefficient of +.59 was obtained--not only a statistically significant finding, but of remarkably large magnitude. While both measures are based on preferences in response to visual stimuli, the Graves Test has nothing whatever to do with hue or saturation, and no very simple relation to brightness, and the responses of the subject to it are not evaluated by reference to a consensus of his fellow subjects. In a further study with 20 of the same subjects (1955b), Granger measured degree of agreement with the consensus in preferences for color *combinations*, and found it to be correlated with the Graves test to the extent of a coefficient of +.73. Less satisfactory evidence of an even greater generality of measures of this sort is presented in an earlier paper by Eysenck (1940 a). His primary measure was of degree of agreement with the consensus in preference rankings of materials drawn from the visual arts--*e.g.*, sets of statues, vases, clocks, etc. This measure was correlated with measures of the same kind based on reactions to odorants, to colors, and to polygons, to the extent indicated by coefficients of +.75, +.53, and +.68. These correlations are based upon only 8, 12, and 14 cases respectively, but two of them are so large as to be statistically significant despite the small number of cases.

Such generality for degree of agreement with the consensus in sensory preferences, a generality extending even to agreement with the consensus in preferences among works of art, would be so important if valid, that it is surprising these two pieces of evidence have not been followed up. Midway in some research on reactions to works of visual art (Child, 1962), I became aware of the importance of testing these claims, and was able to make a beginning in this direction with the second of two groups of 22 subjects used in that research. The stimuli and measures of consensus will first be described, together with an indication of the internal consistency of most of the measures.

1. *Odorants.* 60 diverse odorants were prepared in small brown bottles. They varied from high-quality perfumes through assorted materials of household and medicinal use down to asafoetida. The 60 odorants were divided at random into two sets. A subject first sorted Set A into 10 groups of 3 each, from most liked to least liked. He then did the same for Set B. Separately for each set, inter-person correlations were obtained and factor-analyzed by the centroid method. Personal loadings on the first factor for Set A were correlated +.60 with the corresponding measure derived independently from Set B. The measure of agreement with consensus which will be used is the mean of these two loadings.

2. Hues. Two sets of Munsell color chips were separately ranked for preference: (1) at value 5 and chroma 2, the 10 hues labeled, in the Munsell system, 5 R, 5 YR, 5 Y, etc.; (2) at value 7 and chroma 4, the same 10 hues. The measure of agreement with consensus which will be used consists of first-factor loadings based on inter-person correlations for all 20 stimuli. An approximate measure of internal consistency was obtained by determining, for each set of 7 stimuli, the agreement between each person's rankings and the mean rankings, and correlating the two indices of agreement. The correlation is +.70.

3. Chroma (i.e., saturation). Two sets of Munsell color chips were separately ranked: (1) for the hue 7.5 PB and value 3, chips at chroma 2, 4, 6, 8, 10, 12, and 14; (2) for the same hue but value 5, the same seven chroma. The calculations exactly paralleled those for hues, and the correlation for internal consistency was +.71.

4. Values (i.e., brightness). Again, two sets of chips were separately ranked: (1) for the hue 5 P and chroma 2, values 2, 3, 4, 5, 6, 7, and 8; (2) for the hue 5 R and chroma 2, the same seven values. These rankings, like the others with Munsell chips, were made by moving the chips around on a large sheet of neutral gray paper of value approximately 5. The calculations again paralleled those for hues, but the correlation for internal consistency in this instance was negligible, -.16.

5. Colors. This ranking was obtained incidentally in the course of administering a test for color vs. form dominance modeled after one described by Eysenck (1941). The colors were 10 rectangular pieces of cardboard obtained at an art supply store, differing most obviously in hue but also in value and chroma (though all were of rather high chroma). The measure of agreement with consensus was correlation of individual ranking with mean ranking. In view of the small number of items, no split-half correlation was computed.

6. Polygons. This ranking too was obtained incidentally as part of a color-form test. The stimuli were 10 polygons chosen, with an eye to diversity, from those presented by Birkhoff (1933), and cut out of white cardboard. The measure of agreement was provided by first-factor saturations emerging from a factor analysis of interpersonal correlations. Again, no measure of internal consistency was computed.

7. Paintings. The stimuli here were 12 sets of paintings, each set consisting of 60 postcard-size reproductions, described in Child (1962). The measure of agreement with consensus was the average of the 12 first-factor saturations. The split-half reliability coefficient was +.80.

On the same group of 22 subjects, three other measures of esthetic sensitivity are available--measures in which the subject's responses are compared to some standard of value other than group consensus. These other measures were as follows:

8. The Graves Test, described earlier in this article. It was administered individually, using the test booklet rather than slides, but to save time only one-third of the items were used (approximately a random selection of one out of every successive three items in the test, but using only items presenting two rather than three designs). This abbreviated Graves Test had a split-half reliability of +.60.

9. The Bulley Test of esthetic judgment, also described earlier. The test was administered from the illustrations in Bulley's book (1951), using only the 19 items which, in the study reported earlier in this paper, had shown the highest correlation with original total score. This abbreviated test had a split-half reliability of +.54.

10. Paintings measure of esthetic preference--the correlation between an individual's preference ratings of paintings and experts' ratings of the esthetic quality of the paintings. The paintings and responses to them were the same ones referred to above in connection with the 7th consensus measure. The expert ratings are described in Child (1962). The split-half reliability of this measure was +.90.

The correlations among all these variables are shown in Table 2.[5] I would summarize and interpret as follows the findings that relate to the present inquiry:

1. There is a suggestion of generality in the measures of agreement with consensus in preference ratings of simple visual stimuli (variables 2-6). The results for variable 4 (Values) constitute a marked exception, but otherwise the indication of positive correlation is consistent. This evidence provides some independent confirmation of this portion of the findings in Granger's much more thorough study (1955a). For odorants, the split-half correlation reported earlier provides excellent evidence of generality within that sensory domain, though for stimuli rather more complex than the simple visual stimuli.

2. There is no suggestion whatever that extent of agreement with consensus in the preference rating of odorants (variable 1) is positively related to corresponding measures obtained with simple visual stimuli (variables 2-6) or with paintings (variable 7). In view of the larger number of cases, and greater variety of visual stimuli, used, here, this finding casts doubt on the finding reported by Eysenck with a very small number of cases. There is

---

[5] Correlations were computed for two additional variables which are omitted for lack of significant correlations either for internal consistency or with other variables. One was a consensus measure for poem preferences; the poems were ones selected and used by Eysenck in a study (1940b) in which the intent was to choose poems fairly homogeneous in esthetic value, in order to minimize a general factor in the inter-person correlation matrix. The other was an abbreviated form of the Meier Test (1940); I abbreviated it by using only the double-scored items, and unfortunately the responses of the present subjects to these items were largely invariant.

TABLE 2

INTERCORRELATIONS OF VARIABLES FOR 22 SUBJECTS IN PREFERENCE STUDY

VARIABLES	VARIABLES (numbered as in left-hand column and in text)								
MEASURES OF AGREEMENT WITH CONSENSUS	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Odorants	-.44*	+.11	-.16	-.14	-.31	+.05	+.52*	-.08	-.23
2. Hues		+.25	+.20	+.24	+.29	+.05	-.22	-.16	-.01
3. Chroma			-.51*	+.40	+.26	+.25	+.03	-.45*	-.25
4. Values				-.43*	-.15	-.32	-.04	+.39	+.24
5. Colors					+.63**	+.04	-.02	-.10	+.12
6. Polygons						+.05	-.31	-.24	-.06
7. Paintings							-.07	-.55**	-.79**
MEASURES OF AGREEMENT WITH EXTERNAL STANDARD OF QUALITY									
8. Graves								+.03	+.01
9. Bulley									+.85**
10. Paintings									

---

* $p<.05$

** $p<.01$

also no suggestion that the odorant measure is related to measures of esthetic judgment (variable 9) or esthetic preference (variable 10) based on agreement with expert evaluation of works of visual art. There is a significant positive correlation between the odorant measure and the Graves Test; in the context provided by the other correlations in the table, I would interpret this as probably a result of sampling error.

3. There is no suggestion whatever that extent of agreement with the consensus in rating simple visual stimuli (variables 2-6) has a consistent relation to measures of any kind based on reactions to works of visual art (variables 7, 9, and 10). The results also fail to confirm Granger's finding of a positive relation between the simple visual-stimulus measures and the Graves Test (variable 8). Since Granger's finding is based on more reliable measures and a larger number of subjects, the weight of direct evidence thus far is still in favor of a positive relation in this one instance. That this relation, if it exists, has anything to do with esthetic sensitivity in response to works of art is now highly doubtful. In the previous section, the Graves test was reported to have a near-zero correlation with a measure of esthetic sensitivity in response to works of art. With completely different subjects, this finding is now confirmed by the correlation of variable 8 with variables 9 and 10 (or, with a consensus standard of esthetic value, variable 7).[6]

4. For reactions to visual art, the consensus measure (variable 7) is negatively rather than positively related to measures of agreement with standards of value derived from expert judgment. This aspect of the findings is dealt with in more detail in Child (1962).

5. For reactions to visual art, a measure of esthetic preference based on agreement of preferences with expert judgment of value (variable 10) is closely related to a measure of esthetic judgment (variable 9), even though the works of art and the source of the external standard, as well as the response called for, differ as between the two measures. This aspect of the findings also is dealt with in more detail in Child (1962).

SUMMARY

Several proposed measures of esthetic sensitivity are examined. The question of whether a test of esthetic sensitivity should yield consistent variation below as well as above a chance value is considered. The Bulley Test, which uses works of art as stimuli, is found to show a closer relation to several pertinent variables than does the Graves Test, which uses abstract designs constructed to exemplify analytic principles of esthetic quality. Of measures of esthetic sensitivty tried here, those which use stimuli other than works of art, and those which assess an individual's reactions by reference to consensus of his peers, show little or no relation to those which use works of art and assess an individual's reactions by reference to an external standard of esthetic value provided by the judgment of experts.

---

[6] The Graves test may, of course, be valuable for other purposes than measuring esthetic sensitivity. Group-difference findings reported in the test manual (Graves, 1948) indicate that it measures something with respect to which art students and non-art students differ greatly.

REFERENCES

Anderson, Rose G. A note on the McAdory and Meier Art Tests in counseling. *Educ. & Psychol. Measurement,* 1951, 11, 81-86.

Barrett, H. O. An examination of certain standardized art tests to determine their relation to classroom achievement and to intelligence. *J. educ. Res.*, 1949, 42, 398-400.

Beebe-Center, J. G. *The psychology of pleasantness and unpleasantness.* New York: Van Nostrand, 1932.

Birkhoff, G. D. *Esthetic measure.* Cambridge: Harvard Univ. Press, 1933.

Bolton, E. B. Brief evaluation of two tests of esthetic judgment. *Peabody J. Educ.*, 1955, 32, 211-222.

Bulley, Margaret H. *Art and everyman*, Vol. I. London: Batsford, 1951.

Carroll, H. A. & Eurich, A. C. Abstract intelligence and art appreciation. *J. educ. Psychol.*, 1932, 23, 214-220.

Child, I. L. Personal preferences as an expression of esthetic sensitivity. *J. Personality,* 1962, 30, 496-512.

Coffman, W. E. The Tests of Developed Abilities: Development of the tests. *College Board Review*, 1957, no. 31, 6-10.

Crannell, C. W. The validity of certain measures of art appreciation in relation to a drawing task. *J. Psychol.*, 1953, 35, 131-142.

Crawford, A. B. & Burnham, P. S. *Forecasting college achievement.* New Haven: Yale Univ. Press, 1946.

Dewar, H. A comparison of tests of artistic appreciation. *Brit. J. educ. Psychol.*, 1938, 8, 29-49.

Dyer, H. S. The tests of Developed Abilities: The tests' background. *College Board Review*, 1957, no. 31, 5-6.

Eysenck, H. J. The general factor in esthetic judgments. *Brit. J. Psychol.,* 1940a,31, 94-102.

Eysenck, H. J. Some factors in the appreciation of poetry, and their relation to temperamental qualities. *Character & Personality,* 1940b, 9, 160-167.

Eysenck, H. J. Sense and nonsense in psychology. Harmondsworth & Baltimore: Penquin Books, 1957.

Eysenck, H. J. 'Type'-factors in esthetic judgments. Brit. J. Psychol., 1941, 31, 262-270.

Granger, G. W. An experimental study of colour preferences. J. gen. Psychol., 1955a, 52, 3-20.

Granger, G. W. An experimental study of colour harmony. J. gen. Psychol., 1955b, 52, 21-35.

Graves, M. The art of color and design. New York: McGraw-Hill, 1941.

Graves, M. Design Judgment Test. New York: Psychological Corporation, 1946.

Graves, M. Design Judgment Test, 1948 Manual. New York: Psychological Corporation, 1948.

Lawlor, Monica. Cultural influences on preferences for designs. J. abnorm. soc. Psychol., 1955, 61, 690-692.

McElroy, W. A. Esthetic appreciation in aborigines of Arnhem Land: a comparative experimental study. Oceania, 1952, 23, 81-94.

Meier, N. C. The Meier Art Tests, I. Art Judgment. Iowa City: Bureau of Educational Research and Service, University of Iowa, 1940.

Chapter III

PERSONALITY CORRELATES OF ESTHETIC JUDGMENT

Can psychological research contribute to our understanding of the arts? One way it might do so is by improving our understanding of the psychological functions of artistic enjoyment or appreciation. A step in that direction can be taken by studying the personality characteristics found to accompany variations among people in the adequacy with which they can make esthetic judgments.

The present paper reports an investigation into the correlates of individual variations in esthetic judgment in visual art among male college students of the present day.

SUBJECTS AND PROCEDURE

Subjects

The subjects of this study were young men, undergraduates in Yale College (except for one male graduate student who was within the age range of the undergraduates). The students were mostly freshmen and sophomores, but a few were juniors and seniors. The most important element of selection for present purposes, apart from the fact that all were students at a particular university, is the fact that most subjects enlisted for the research without knowledge that their reactions to art were to be investigated, so that selection is for the most part without bias in relation to amount of interest in or knowledge of art. The students were obtained as paid subjects in psychological research, described as being concerned with personality and preferences. Solicitation was made through a bulletin board used for securing subjects for experiments done in connection with elementary psychology courses, and through the student employment bureau. Most of the students appeared to have signed up because of an interest in earning money (they were promised the opportunity of at least seven hours of work) or because of an interest in taking part in psychological studies, or both. Since the research continued over some period of time, some students did sign up because of having learned from others that the research was concerned with reactions to art, so that the sample is probably somewhat biased in the direction of including an above-normal number of students with a definite interest in art.

General Procedure

In signing up for the study, each subject made an appointment for the first of two evening sessions held in groups. This first session, three hours in length, was followed by a second group session, generally one week later. In signing up, each student also committed himself to taking part in a single hour's individual session at a later time, ordinarily within a week or two after the second group session. In the group sessions the students took various tests and questionnaires, and these were supplemented in the individual session by some measures which would be obtained only with one person at a time.

At the beginning of the first group session the students were informed that the research was especially concerned with reactions to art, and with aptitudes, skills and other characteristics that might be related to those reactions. They were informed that the research was concerned with generalizations about variables, not about individuals, and that all the information obtained about them would be used in a manner which would guarantee anonymity.

The measures used for esthetic judgment will be described here. The measures used for other characteristics will be described briefly in connection with presentation of results, with reference to sources of fuller information where appropriate.

The Measures of Esthetic Judgment

The measures of esthetic judgment were based on the method used by the English esthetician Margaret Bulley (1951). In preliminary work reported in the previous chapter, the author had concluded that this was likely to be the most satisfactory kind of measure of esthetic judgment in the visual arts that could easily be made available. The method as used here consists of presenting pairs of works of art, in the form of projected slides, with the subject instructed to attempt to judge "which of the two works of art is better esthetically--that is, is the better work of art." The subject is instructed to express an esthetic judgment of his own if he has one; if he has no such judgment, either for particular pairs or in general, then he is to take as his task that of trying to judge which of the two works of art is the one which a group of experts has agreed is the better.

The pairs of slides used in this study were prepared especially for this purpose. A number of assistants, mostly graduate students in art or history of art, or older people with a similar background, selected the pairs. They were instructed to try to form pairs of works of art, the two works to be of the same subject matter or type, and if possible also of the same style, but yet to differ decidedly in esthetic value, in the opinion of the selector. In an effort to prevent the initial pairs from representing the esthetic judgment only of one or a few people, diversity was sought in the characteristics of the judges, within the limits necessarily imposed by the fact that the pairs were all obtained in a brief period in a single community. Pairs were selected by 24 people all together, including people from widely scattered places of origin. However, about a third of the pairs were selected by a single assistant who was unusually successful at the task and worked full time.

All of the pairs that were selected and photographed as slides were then shown to 14 different judges, people who as experts were asked to indicate which work in each pair they considered to be the better. These judges were mostly students in the School of Art at Yale; some were graduate students in history of art, and some were older people in the New Haven area with a similar background. Their judgments were made quite rapidly; each pair was projected on a screen for only about 20 seconds, and the judgment had to be recorded within this time.

Two categories of pairs are used in the present research.

Easier test of esthetic judgment. Of about 3360 pairs selected and judged, there were about 1270 on which either the 14 judges were unanimous or only one or two judges disagreed with the original selector. Out of these pairs 120 were chosen at random to constitute a relatively easy test of esthetic judgment.

Difficult test of esthetic judgment. From the pairs on which there was a lesser degree of agreement among the expert judges, about 1020 were selected as still likely in the long run to yield majority expert agreement with the original selector. The choice was based on two facts about each pair: (1) the number of judges agreeing with the original selector; (2) whether the disagreements occurred more among those judges who most regularly agreed with the original selectors or among those judges who were most prone to disagree with the original selectors. Out of these pairs another 120 were used as a difficult test of esthetic judgment. Here, there would undoubtedly be more pairs than in the easier test for which the expectation of majority agreement in a larger group of experts, were it tested, would fail to be confirmed.

The easier test of esthetic judgment was given in the group session to all of the 138 subjects who took part in this study. From among the first 59 subjects, 21 who took part in a later experiment were given the difficult test in a later individual session; since this, however, is a selected group -- selected for their original scores on the easier test of esthetic judgment -- their results will not be dealt with in this chapter. The last 79 subjects to take part in the study were given the difficult test of esthetic judgment as a part of the group sessions -- for them it was included as one of the things done in the second group session -- and results obtained in this way will be presented here.

The easier test of esthetic judgment as administered in the group session has an alpha coefficient of .87; this is an estimate of the mean of all possible split-half reliability coefficients (see Cronbach, 1951). The difficult test of esthetic judgment as administered in the same way has a much lower alpha coefficient, only .45. For the 79 subjects who took both these tests in the group session the correlation between the two is +.75. These results indicate that the two tests maybe measuring substantially the same personal characteristic. Relation of other variables to these tests will be reported separately, however. Main confidence should be placed in results with the easy test, both because it is a more reliable measure and because it is available for a larger sample.

Except where otherwise indicated, the number of cases on which correlations are based is 138 for correlations of the easy test of esthetic judgment with other variables, and 79 for correlations of the difficult test with other variables. Since, however, the introduction of the difficult test and addition of other measures required dropping some of the procedures midway in the research, smaller numbers are available for certain correlations. Wherever this is true, the number will be provided in the text or table.

RESULTS

A. Background in Art

Assessment of background in art was made through an interview conducted at the end of the individual session with each subject. For the first 59 subjects, the interview was somewhat variable because several interviewers were involved and less attention was given to standardizing the other procedures which occupied most of the individual session. The last 79 subjects were all interviewed by one person who followed a uniform plan of inquiry. For all subjects, the interview was recorded in writing, as nearly verbatim as possible, while it was being conducted.

Four variables in the background were subsequently rated by two judges working independently. All four variables could be rated for all of the last 79 subjects and for most of the earlier ones. Table 1 shows the intercorrelations of these variables (each measured by pooled ratings of the two judges), and in the diagonal appears the inter-judge correlation for each variable. The number of subjects for whom each variable was rated is given in the right-hand column.

The correlation of each of these variables with the two measures of esthetic judgment, the easy test and the difficult test, is also given in Table 1. The correlations with the difficult test are based on only 79 persons; the correlations with the easy test are based on all persons for whom the particular rating is available, as indicated in the last column of the table.

The correlations with esthetic judgment are all positive. This result is of course consistent with the obvious hypothesis which we may confidently expect to be true, that esthetic judgment is partly dependent upon knowledge about art and about other people's evaluations of art. The magnitude of these correlations does not, however, give us any basis for estimating to what degree esthetic judgment is dependent on specific experience with art and to what extent on more general influences. For what our subjects said in the interviews

Table 1

Intercorrelations of Background Variables (in the diagonal, inter-judge correlations corrected by the Spearman-Brown Formula)and Relation to Esthetic Judgment

	1.	2.	3.	4.	Easy Test	Difficult Test	N
1. Education in Art	.89	.57	.42	.22	.49	.39	137
2. Experience in Galleries, etc.		.89	.46	.36	.49	.36	137
3. Art-Related Hobbies			.86	.26	.21	.08	126
4. Family Attitude				.88	.18	.25	132
The 4 background variables in a multiple regression equation					.59	.44	

makes very clear that background in art is obtained partly through a person's own initiative and is thus a product of, as well as an influence on, his esthetic sensitivity and interest in art. At the extreme, for example, one student reported that he seeks out art galleries in any city he visits, while another said that he had never been to an art gallery because there are none near where he lives, a suburb of New York City.

Control of background in art as a variable would be desirable. Wherever a significant correlation is found between esthetic judgment and another variable, it is possible that the relationship may be entirely explicable by a common relation of each of these to background in art. In every instance of a significant finding, therefore, the effect of this common relation will be controlled statistically by the technique of partial correlation. This control might, however, obscure a genuine direct relationship between esthetic judgment and the other variable, for the common relationship of these two to background in art may in part result from what they have directly in common, rather than being the entire reason for their having anything in common. Hence, priority in presentation will be given to the simple correlation between esthetic judgment and each of the other variables, and the partial correlation (with the four measures of art background controlled) will be presented afterward in all instances where it is relevant because of being appreciably larger or smaller than the simple correlation.

B. Skill in Perception of Visual Form

A second hypothesis was that esthetic judgment in the visual arts would be dependent in part upon the ability to perceive visual form. For this purpose several tests were used which had in common the requirement that visual form be perceived under circumstances which make it somewhat difficult to do so. Five of these tests were derived from tests used by Thurstone (1944) in his study of the factorial relations of tests of visual perception. The first four of these were adapted from Thurstone in an identical manner: to wit, his stimuli were used, but they were projected on a screen at a group session for 15 seconds each, and each subject recorded his response. Each test was begun by showing the first item for a longer period as an example of the task, and the correct answer to it was mentioned incidentally. The score for each test was the number of items for which a correct answer was given.

1. Dotted Outlines

The subject was instructed that the slide would show a pattern of dots which might form a letter or number, and he was to write down that he saw if he saw anything within the time limit. The score was the number of letters or numbers seen correctly. The Dotted Outlines test showed a correlation of +.04 with the easy esthetic judgment test and +.15 with the difficult esthetic judgment test. Neither is statistically significant.

2. Hidden Digits

The subject was instructed that among a number of dots on the screen would be some arranged in a pattern which would make a letter or number. The score correlated +.07 with the easy test of esthetic judgment and -.01 with the difficult test.

3. Mutilated Words

Each stimulus consisted of a typewritten word from which pieces of the letters had been sliced out, and the task was to tell what the word was. This score was correlated +.06 with the easy test of esthetic judgment and +.09 with the difficult test.

4. Picture Perception

These stimuli were drawn by Thurstone from the Street Gestalt Test. A simple drawing of an object had segments sliced out, and the task was to recognize the object despite this mutilation. The score was simply the number of pictures reasonably well identified, with no qualitative score beyond. This score correlated +.22 with the easy test of esthetic judgment and +.15 with the difficult test. The first of these correlations is significant at the 5% level; when the art background variables are statistically controlled, the partial correlation coefficient is +.18, which is only a fraction short of being significant at the 5% level.

Since these first 4 tests appeared likely to be measuring a similar variable and were in fact positively intercorrelated for our data as well as for Thurstone's, they were also pooled to form a single variable. This showed a correlation of +.22 with the easy test of esthetic judgment and +.16 with the difficult test. Here, too, the correlation with the easy test is reduced when art background is statistically controlled; the partial correlation becomes only +.12 and is not significant.

5. Hidden Figures

The material here consisted of Thurstone's version of the Gottschaldt figures, in which the subject's task is to locate and mark a relatively simple geometrical figure embedded in a more complicated pattern. The problems were relatively simple in comparison with the more difficult ones used by several recent investigators (cf. Witkin _et al._, 1962, and Gardner _et al._, 1959). The stimuli were presented in a booklet prepared by photo-offset from the material published in Thurstone's monograph, and his time limits were followed. On several of the sub-tests, some subjects finished long before the time limit was up, while others had still not finished at the end. A speed measure was therefore introduced by having each subject copy at the bottom of the page the time he finished, from a time record kept by the experimenter on the blackboard. We felt that under our conditions of testing the subjects' self-report of time in this fashion should be valid.

Several separate scores were obtained from the Hidden Figures test. The number of items answered correctly and the speed with which the subject completed each part (measured quite without regard for the correctness of his completions) were so closely related that they were combined into a single score, which was insignificantly correlated with esthetic judgment in the wrong direction with relation to the prediction: -.03 with the easy test and -.13 with the difficult test. The number of items answered incorrectly showed a negative correlation, as predicted, with both tests of esthetic judgment, -.14 with the easy test and -.15 with the difficult test. When art background is statistically controlled, the former correlation rises to the significant value of -.20, while the latter correlation remains -.15. Finally, an analysis was made of whether or not the subject made any of the odd or grotesquely wrong responses that some subjects were found to make. Making these odd responses was correlated negatively, as should be expected, -.07 with the easy test and -.09 with the difficult test, but these values are much too small to be significant.

6. Spatial Visualization

The final test which might be regarded as related to skill in perception of visual form was not derived from the Thurstone materials although it was similar to some of his tests. It was a test of spatial visualization; the one chosen for use was the Psychological Corporation Test of Spatial Relations. Though manipulation of visual imagery is what appears to be called for by the test, the fact that 3-dimensional understanding of 2-dimensional representation is required justifies its being considered for present purposes under the heading of perception of visual form. This test, ordinarily given with a 30-minute time limit, was given here with a 15-minute time limit, and thus became more of a speed test than it ordinarily would be. With this restricted time limit, only one or two subjects were able to finish the test. One score based on it was the number of items answered correctly, which showed practically no tendency toward relation with either test of esthetic judgment: -.01 with the easy test and +.08 with the difficult test. A second score consisted of the number of items answered incorrectly; this score showed insignificant correlations of -.14 with the easy test of esthetic judgment and -.05 with the difficult test. Control of the art background variables raised the first of these to the significant value of -.195, which is in the predicted direction.

test. A second score consisted of the number of items answered

In general, then, the results do not strongly confirm the expectation that esthetic judgment would be related to skill in perception of visual form. The correlations obtained do show a definite tendency to be in the predicted direction, and some of the correlations meet the criterion of statistical significance at the 5% level. From these facts one might conclude tentatively that there is a very slight tendency towards the correlation of esthetic judgment with skill in perception of visual form, but there is no indication of any very important relationship.

C. Skill in Perceiving Human Meaning in Ambiguous Stimuli

A third hypothesis was that esthetic judgment in visual arts might depend partly upon skill in finding human meaning in the non-representational elements in art whose integration with the representational elements seems to play an important part in determining the esthetic value of paintings.

One test employed for this purpose was Barron's Inkblot Test (Barron, 1955), intended to measure the individual's tendency to make human movement responses to ambiguous stimuli and thus to serve one measurement function of the Rorschach Test in a more objective and reliable way. The stimuli used by Barron were employed, but they were projected on a screen in a group situation instead of being administered individually. Each stimulus was shown for between 20 and 25 seconds, and the subject was instructed to write down the first thing he saw. Preliminary explanation indicated that different people see different things, and one person may see different things at different times, but that what was wanted was the first thing each person saw. The only indication given of what kind of percept was of interest to the investigator was the statement that like other tests just presented it was an attempt to measure originality. The inkblots were presented in the order of their likelihood of evoking human movement responses as reported by Barron for his standardization group. Responses were scored twice, once by a lenient standard and once by a stricter standard for deciding whether a movement response had been made. The scoring by the more lenient standard showed a correlation of +.13 with the easy test of esthetic judgment but only +.04 with the difficult test. The score derived by the use of a stricter criterion was correlated +.17 with the easy test and +.06 with the difficult test. These correlations are all in the expected direction, but none are significant and they do, at any rate, not suggest any very important relationship.

Attempts were also made to measure the accuracy with which a person could perceive appropriate connotative meaning in visual stimuli such as appear in art. For this purpose we used works of art, but abstract art in order to eliminate representational meaning. There were no tests already developed; therefore the tests used were tried out in the course of the research itself.

Attempts were also made to measure the accuracy with which a person could perceive appropriate connotative meaning in visual

First of all, an attempt was made to measure the extent to which a person could judge appropriate emotional connotations in the use of color and form in abstract paintings. Slides were obtained of 10 rather diverse abstract paintings. The subject was supplied with copies of the list of mood words prepared by Hevner (1937) for her various studies of connotative meaning. Each painting was projected on the screen for about 40 seconds, and the subject was asked to check all the mood words he felt to be appropriate in describing what the painting seemed to embody or express. These judgments were scored in two ways. The first score was the number of responses the individual made which were common responses as judged by the criterion of their having been made by as many as 20 out of the 59 subjects to whom this test was applied. The second score was the number of unusual responses he made, as determined by the criterion that the response had been made by no more than 5 out of the same group. The reasoning which underlay these two scores involved taking the characteristic responses from this group of subjects as a tentative index of perceptive or appropriate response. Thus, for a person to indicate the same emotional meaning found by a sizable number of his fellows might imply accuracy in such perception, and for him to indicate emotional meaning found by almost no one else might suggest a careless or imperceptive inaccuracy in perception of emotional meaning. (Obviously, this is a fallible criterion for which it might be better to substitute the judgments of people of known competence in observing paintings with understanding.) The scores as thus derived showed a relationship with the easy esthetic judgment test which was not in the direction expected. The number of common responses made by a subject was correlated -.31 ($p < .02$). The relation between this measure and the easy esthetic judgment test falls to the insignificant value of -.13 when art background variables are controlled. The number of unusual responses simply showed no appreciable correlation, -.04. A major difficulty in determining the meaning of the significant finding with number of common responses is a consideration which might well have led to not using this particular measure in the first place: the fact that this score is greatly influenced by the sheer numerousness of responses a person makes, since in this test he is invited to check off as many mood names as he finds appropriate, and people do in fact vary widely in their number of responses.

In the remainder of the research, therefore, we substituted another test to eliminate this difficulty. Sixty pairs of abstract paintings were selected in advance by two judges, with the attempt to make each pair differ in connotative meaning along the three dimensions which Osgood _et al._ (1957) have stressed--pleasantness vs. unpleasantness, activity vs. passivity, and strength vs. weakness. For those items on which a majority of the subjects agreed with the judges who originally picked out the paintings the scoring was based upon the combined opinion of the original judges and the majority. For those items on which there was not such clear agreement, or there was even disagreement between the original judges and the majority of subjects, scoring was done simply by the criterion of agreement with the original judges (because of the possibility that the original judges, more familiar with visual art, might be displaying greater wisdom than the majority of subjects).

A separate score was obtained for each of these two sets of items, but the results were of negligible magnitude. Connotative-meaning score on the first set of ideas was correlated +.11 with easy esthetic judgment and .00 with difficult esthetic judgment. Connotative-meaning score on the second set of items had corresponding correlations of -.17 and -.19. None of these results provide significant evidence of a relationship.

An attempt was also made to measure the accuracy with which a person could perceive masculinity vs. femininity of connotative meaning in abstract painting. The judges pre-selected pairs of abstract paintings or simple designs with the intention of having one feminine and one masculine in connotative meaning, or of having one contain elements of both sexual meanings and the second be masculine or feminine alone. The subjects were then told how the pairs had been picked out and asked to judge the sexual connotative meanings of the two paintings of the pair, with the proviso that the two be placed in different ones of the three categories: masculine, feminine, and mixed. Here only those items were scored for which a majority of the subjects showed agreement with the original judges. This measure was not available on many subjects who took the difficult test, but was available on subjects who took the easy test, and the correlation with esthetic judgment was close to zero.

Clearly, then, we fail to find any consistent evidence to support the hypothesis that esthetic judgment in visual arts depends upon skill in finding human meaning in the non-representational elements in art. The hypothesis may, of course, be correct, but if so, the relevant skill is much more specific than might have been imagined and is not measured by the testing instruments used here.

D. Behavioral Measures of Cognitive Style

A number of discussions of cognitive style in recent psychological publications suggest dimensions to which esthetic judgment might well be related. This relation will be developed further in the following section which deals with questionnaires intended to measure comparable dimensions.

Concepts of cognitive style have been developed particularly in connection with behavioral measures and the interrelationships found among them. For the most part the magnitude of correlation found among these measures in previous research is rather low, so low as to suggest that the very restricted measures that have been principally used may be ineffective as measures of general personality characteristics. For this reason we did not invest a great deal of time in their measurement, at the probable expense of making it more difficult to obtain subjects, but included only a sampling of behavioral measures that could be administered in a relatively short time. An exception was made for measures dealt with in the preceding sections, some of which are among the behavioral measures which have been considered by other writers to be relevant to cognitive style. The *a priori* argument for their relevance to esthetic judgment seems stronger than for a variety of other measures, and hence more testing time was devoted to them.

1. Measures Derived from the Color-Naming and Color-Word Test

This test, devised by Stroop, (1935) has been used by Thurstone and by later investigators (Broverman and Lazarus, 1958; Gardner _et al._, 1959). As used here, the stimuli were projected from slides, but in an individual testing situation. First is projected a 10-line list of repetitions of the words red, blue, green, and yellow, after the subject has been shown a sample line and instructed to read the words aloud as rapidly as possible. Next is projected, again after a sample line, groups of asterisks equivalent in number and spacing to the color words, with each group of asterisks painted in one of the four colors, red, blue, green, and yellow. The subject this time is to call out the names of the colors as rapidly as possible. The ratio of the time required for naming the colors to the time required for reading the color words (more precisely, the difference between the logarithms of the two times) has been used as a possible measure of relative dominance of concrete vs. conceptual modes of thought (Broverman and Lazarus, 1958). This measure was negligibly related to the scores on esthetic judgment (correlations were +.07 with the easy test and -.10 with the difficult test).

The subject is next shown a list of the identical color names, this time colored in the same order in which the asterisk groups were colored, but with the color always being one of the three colors other than the one named in the word. His job now is to call out the color in which the word appears, disregarding the word itself. This is a difficult task for most subjects, and performance is greatly slowed. The measure was the ratio of the length of time required for calling out the 100 colors under these circumstances to the time required for calling them out in the previous task in which there was no element of conflict. (Again the measure actually used was the difference between the logarithms of the two times.) This measure was used on the assumption that non-disruption by this conflict might be an index of a general ability to tolerate simultaneous presence of conflicting stimuli, an element which might be important in the ability to understand complex works of visual art. No such correlation was found as would be predicted on this basis. The conflict score was negligibly correlated with the easy esthetic judgment measure, +.03; too few cases were available for testing the correlation with the difficult esthetic judgment test.

Finally, another slide was shown which was like the last one except that the words had no relevance to colors. The subject was again asked to call out the colors as rapidly as possible. Immediately after he had finished, the projector was turned off and he was then asked whether he could perhaps recall any of the words that had appeared even though he had been asked to concentrate on the colors in which they appeared. If he did, he was asked to give the words. The number of words thus remembered incidentally when instructed to attend to other things was used as a possible measure of scanning tendency--that is, tendency to notice a variety of stimuli beyond those on which attention is especially focused. High score on this measure might indicate a consistent perceptual tendency which would favor ability to comprehend the complexities of visual art. The correlations found were negligible and in the wrong directions, -.04 for the easy esthetic judgment test and -.08 for the difficult esthetic judgment test.

2. Measure Obtained with Two-Hand-Coordination Test

Thurstone described this test on pages 49 to 51 of his monograph. The subject is required first to perform a repeated series of movements with his dominant hand, then to do separately a different series of related movements with his non-dominant hand, and finally to carry out the two series of movements simultaneously. This last task produces a great deal of conflict for most subjects. Ability to carry out the two series of responses simultaneously with anything approaching the speed of the single performance of one of them might indicate a high degree of toleration for simultaneous attention to conflicting external stimuli or to internal responses to them. The intercorrelations obtained by Thurstone with this test suggest indeed that it is possibly a useful measure of such a general ability. In his study it is the measure with the highest loading on a factor which he believed to represent "the manipulation of two configurations simultaneously or in succession." The prediction that such a measure might therefore be related to esthetic judgment was not fulfilled. There was even a tendency in the opposite direction, a correlation of +.19 for the easy esthetic judgment test and +.12 for the difficult test with the extent to which discoordination was displayed by the subjects on this test. Neither of these coefficients reaches a significant level as N is only 89 for the easy test and 31 for the difficult test. When art background is partialed out, however, the correlation with the easy test rises to the highly significant value of +.27, while that with the difficult test retains approximately the same value (+.11).

3. Measures of Perceptual Constancy

The ability to maintain perceptual constancy--that is, to perceive as unvarying an aspect of an object whose situation or context causes its apparent aspects to alter--might be thought of as a variable in perceptual functioning which would be important in understanding visual art. Only a single kind of constancy was measured, shape constancy. In a procedure based on that used by Thurstone the subject looked through a slit in a board erected at one end of a long table and was asked to judge the true shape of a diamond-shaped cardboard laid flat on the table at the far end, some 7 feet away from him. Two separate observations were made, one with each of two diamonds; the intercorrelation between the two was +.75, and the pooled score was used. The correlations with esthetic judgment were very small and tended in the opposite direction from that predicted. With the easy esthetic judgment test it showed an insignificant correlation of -.16 and with the difficult test, -.10.

4. Color vs. Form Dominance

It seemed reasonable that appreciation of visual art might depend partly on ability to look at objects with careful attention to color as well as to form. Both the history of psychological testing and casual observations of reactions to art suggest that many people are in a sense afraid of attending to color, and that attention to form alone is a possible retreat from emotion-arousing values of visual experience. In an attempt to measure this characteristic we drew on the variety of tests of color-vs. form-responsiveness which had been assembled from various sources by Keehn (1955). From the many tests he employed we selected six on the grounds of their correlation with other color-vs.-form tests in Keehn's results, brevity of time required for their administration, and ready possibility of preparing the materials required. These tests were the ones he called grouping, common elements, memory for letters, card sorting, odd-one-out, and stepping stones. These six tests did indeed show a tendency toward positive intercorrelation, though without perfect consistency. If we regard them as a single 6-item test, the alpha coefficient of internal consistency has the value of .50. The personal characteristic measured by this battery of six items has virtually no relation to esthetic judgment, as its correlation with the easy test is +.07 and with the difficult test -.05.

5. Measure of Non-Constancy

"Perceptual constancy" refers to the tendency to perceive things as the same despite changes--to see a playing card, for example, as rectangular despite the angle at which it is viewed, or to see a necktie as the same color despite variations in the light falling on it. Ability to get away from perceptual constancy when instructed to do so is a possible measure of a flexibility that might be related to adequacy of understanding of visual works of art. A single measure was used. With the same apparatus as for the constancy measure the subject was shown a square held upright at a comfortable viewing distance from the slit through which he was looking. While he watched, this square was then laid down at the far end of the table, and he was asked to judge the shape of the diamond that the square now made as an apparent shape in the visual field. The judgment was made, as in the constancy test, by having the subject choose that one of a number of forms which seemed appropriate to him from a set of forms modeled after the ones used by Thurstone. This measure showed a negligible correlation of +.10 with the easy test of esthetic judgment and a slightly more suggestive correlation of +.20 with the difficult test, not quite significant at the 5% level; both these correlations are in the predicted direction.

6. Measures Obtained with Aniseikonic Lenses

Recent investigations (Gardner, *et al.*, 1959) have found evidence that reactions to wearing aniseikonic lenses (which create inequalities of size and shape in different parts of the visual field) may reflect individual differences in general tolerance for novel or unrealistic experience. The particular lenses we used were ones which for a person tolerant of the novel experience create a gradient in size across the visual field from left to right, which may be perceived either as a gradient in size among objects which retain their normal apparent distance or as a gradient in distance among objects which retain their normal apparent size. The subject was seated in a standard place in the experimental room where the location of objects was kept fairly although not absolutely standardized. He was instructed to put on the lenses and then report what he saw, having been forewarned that they lead some people to see things rather differently from usual, either immediately or after a while, but that some people continue to see things quite the same as usual. The experimenter noted the subject's comments and the time when they were made. Then about a minute and a half after the subject had put the lenses on, the experimenter instructed him to hold his hands up in front of him but not immediately next to each other and to report anything further that he observed. In a subsequent half-minute or so the experimenter noted any further spontaneous comments and if necessary asked the subject how things looked now.

For the first 59 subjects an over-all rating was made of the extent to which the individual showed a recognition and acceptance of the perceptual changes which the lenses might produce. This measure had a highly significant correlation of +.40 with the easy test of esthetic judgment. However, the various aspects of response to the lenses seemed somewhat separable; and in dealing with the data obtained from the entire body of 138 subjects we therefore decided to attempt a more detailed analysis of the response. Accordingly, four separate ratings were made. First rated was the extent to which the subject reported novelty in perception during the initial period before his attention was called to his hands. Response varied all the way from the very common effect of seeing nothing unusual to seeing the room and furniture as markedly distorted. The second was a rating on the subject's response to looking at his hands. Although seeing one hand as decidedly larger than the other was the most common response, variation was found all the way down to seeing no difference from their usual appearance. The third rating was of the maximum distortion reported in the appearance of the room and furniture, whether this maximum was reached before looking at one's hands or only in the period afterward. Finally, a fourth rating was made, possible only for those people who did report some decided distortion of the physical environment; it was a rating of the direction in which the effect was reported. This was made initially because of the possibility that subjects varied from seeing an appropriate distortion all the way to seeing an opposite distortion as a form of struggle against the effects produced by the lenses.

In the course of making the ratings, however, it became apparent that this did not seem to be the nature of the variation; that it was rather a variation in the extent to which people describe the effect in terms of the changed appearance of a single relatively isolated element in the visual field, such as the table in front of and below their eyes, as against reporting the effect in terms of the total visual field. Thus, consider the slope of the table in front of and below the subject's eyes: this might be reported as sloping up to the right, the view that should be obtained if this were the only distinguishable object within the visual field, since the fact that the parts of it to the right are larger than the corresponding parts to the left should give an impression of their being closer to the eyes and hence higher in position. The table top might, however, instead be reported as sloping down to the right. This might be thought of as a consequence of seeing the table as a part of an entire room in which the left-hand side is stretched off far into the distance with the middle of the room therefore being higher than normal and horizontal objects below the middle stretching out towards that higher point.

Each of the first three measures, which could be applied to all of the subjects, confirmed the finding found with the first group alone, the sole exception being one measure with the easy test. The effect obtained while looking at the hands showed a correlation of only +.13 with the easy esthetic judgment test, but a very sizable correlation of +.24 with the difficult esthetic judgment test. The effect on perception of the physical environment up to the point of looking at the hands was correlated +.22 with the easy test and +.25 with the difficult test. The maximum effect obtained in perception of the physical environment was correlated +.31 with the easy test and +.33 with the difficult test. Adding together the ratings on these three variables, the correlation of this total score with the easy test was +.29 and the correlation with the difficult test was +.32. These correlations are independent of relations to measures of art background; the partial correlation with easy esthetic judgment score is +.32 and that with the difficult score is +.37. These correlations are all highly significant statistically. The other rating, applicable only to those who showed substantial distortion, showed a correlation of +.41 with the easy esthetic judgment test and +.12 with the difficult test, both in the direction of a positive correlation between the direction of effect that seems dependent upon perception of the total visual field and goodness of esthetic judgment. The correlation of the single rating with the easy test is remarkably high and is significant at the 1% level even though based on only 58 cases, but confidence in its meaning must be reduced by the failure to find any sizable confirmation with the difficult test. The partial correlation of the easy test with this single rating, with art background controlled, is only +.23 still significant but much lower.

For the general ability to accept and report a distortion of the visual field, then, the original finding of a decided correlation with esthetic judgment is markedly sustained. This stands out alone among the behavioral measures as showing a marked correlation with esthetic judgment. One might therefore be inclined to question the reality of the effect, supposing it to be rather the one significant result that arises at random out of no true effect. However, the effect was first observed among the first 59 cases and then confirmed separately as a statistically significant finding for the following 79 cases, not only for the easy esthetic judgment test but now also for the difficult esthetic judgment test. This provides rather convincing evidence that there is a true relationship here.

7. Response-Set Measures Obtained from Questionnaires

Of late the possibility has been suggested that various sets adopted by a person in answering a questionnaire may be useful indices of more general personality characteristics. They might well be indicative of rather general characteristics of cognitive functioning. This has been disputed by some writers, and the evidence is far from clear. Questionnaire measures whose assumed meaning depends upon the content of the statements to which the subject responds will be considered in later sections. But most of the questionnaires dealt with there could also be scored for consistent response tendencies with a meaning independent of the content of the statements, and these behavioral or "response-set" measures will be considered here.

One such measure was agreement set--that is, the extent to which the person tends to agree with or accept as characterizing himself statements which appear in questionnaires, where the statements equally often represent one and the other end of the dimensions being measured, so that one would expect on the whole that genuine reflection of personality characteristics would not lead to consistent deviation in either direction from agreement with 50% of the items. There was indeed a rather consistent characteristic measured by the count of number of items agreed with. The split-half correlation was determined for the first 59 subjects and was found to be .80. Exploration of the relation of this variable to the other personality characteristics under study here did not suggest that it was a measure of any general significance. At any rate, it failed to show any decided relation with esthetic judgment. The correlation with the easy test of esthetic judgment was only -.12, although the correlation with the difficult test rose to the value of -.18. With art background controlled statistically, the latter correlation rises to -.23, a value significant at the 5% level. While not impressive, these results are in the direction to suggest that goodness of esthetic judgment tends to go with lack of agreement set. A possible interpretation consistent with other results to be presented later, is that esthetic judgment is correlated with general independence of judgment, and that agreement set indicates lack of such independence.

It is possible, however, that a systematic tendency to disagree with statements, to reject them, might indicate lack of independence of judgment just as much as would a systematic tendency to agree. Hence an additional measure of response set that was determined was systematic tendency to deviate in either direction from 50% agreement. This might be called agreement-disagreement set and be thought of as representing the extent to which the individual tends to have his responses determined by a consistent tendency toward either agreement or disagreement rather than by the content of the items. This measure of response set also shows little indication of relationship to the other personality measures obtained here, but it does show some suggestion of a relation with esthetic judgment. This time the suggested relation is in the opposite direction to what the general interpretation being offered here would suggest. The correlation of this measure is +.16 with each of the two tests.

Most of the questionnaires in this study gave the individual four alternatives in responding to each item. He could say that the item tended decidedly to characterize him (or that he decidedly agreed with it), that the same was true slightly, or he could make either of these responses in the opposite direction of disagreement with or rejection of the statement. It was possible therefore to score all these questionnaires for the extent to which a person tended to make a decisive answer at one of the two extremes rather than a less decisive answer in one of the two middle positions, and this was an even more consistent characteristic of the person's response to the questionnaires than was agreement set. The split-half correlation here for the first 59 subjects was .97. This variable also showed no sign of a relation to esthetic judgment, and indeed there seemed no persuasive reason in advance to suppose that it would. The correlations were +.01 for the easy test and +.04 for the difficult test.

Here too there was the possibility of using a measure which would take account of deviation in either direction, and for such a measure there are grounds for predicting a relation to esthetic judgment. The extreme of answering almost always with an extreme answer or that of answering almost always with a guarded, reserved, or mild answer both are forms of behavior that suggest lack of discrimination, since surely for a questionnaire containing a great variety of items discriminating attention to and response to the items would suggest a decided response to some and a milder or qualified response to others. Thus a deviation in either direction, toward decisiveness or toward non-decisiveness, seems possibly indicative of an undiscriminating form of response. But where is the middle position to be found? While an even division between decisive and guarded answers would provide a maximum of discrimination, there is no reason to regard this division as especially appropriate to the meaning of the terms "decidedly" and "slightly" used in the instructions to subjects. The amount of discrimination to be regarded as optimal may perhaps more reasonably be determined by the responses of the subjects. This amount was set as simply the degree of decisiveness exhibited by the median subject. Furthermore, in order to obtain a measure truly independent of the decisiveness measure itself, the distribution of deviations on either side of this central point was normalized. This measure of deviation toward decisiveness or guardedness, considered in advance as a possible measure of lack of discriminating response, showed no correlation with the easy test of esthetic judgment, the coefficient being -.01. It did, however, show a correlation of -.25 in the expected direction with the difficult test of esthetic judgment (increased to -.26 with statistical control of art background). This result is worth considering for its possible indications--both for the discriminating response-set score and the difficult esthetic judgment score--that the measures used here may genuinely reflect some relatively high level of fineness of cognitive discrimination.

E. Questionnaires on Cognitive Style

Sensitivity to esthetic values seems likely to be related to more general variables of cognitive function. Among likely variables of this sort are several which have emerged in research of recent years, and which may be referred to by the general label of cognitive style. Some have been measured by questionnaires. Some have been measured only by performance tasks. As has already been indicated, the danger of over-specificity and the great time required by performance tasks led to their not being principally relied on in the present research. Instead, questionnaires were used as the main technique in an exploration of the relation of esthetic judgment to general variables of cognitive style.

One published questionnaire, the Myers-Briggs Type Indicator, (Myers, 1962) was used; otherwise the questionnaires are unpublished and mostly devised for this project, although a few of the items were taken or adapted from previous questionnaires. To avoid contamination of variables we were trying to measure separately, items bearing on the arts were avoided in the questionnaires. Within any one the items related to a particular variable were intermingled with others rather than standing as a separate group. All the items were given as statements to which the subject was to respond on a 4-point scale: agreeing with (or accepting) the statement vs. not agreeing with it, and doing so either slightly or moderately. Except as mentioned below, each variable was measured by an equal number of items phrased so that agreement would represent the positive and the negative end of the variable. The score on each variable was a simple count of the number of items on which the subject had responded in a manner indicative of the positive end of the variable.

The internal consistency of the specially-constructed measures has been assessed by a split-half coefficient of correlation, corrected by the Spearman-Brown formula, with each half containing an equal number of items phrased in each direction.

1. Tolerance of Complexity

A first such variable was that of tolerance of complexity. This, along with the next variable to be considered, emerged from considering the concept of tolerance of ambiguity, emphasized in the California Studies of the authoritarian personality (Adorno et al., 1950), and the concept of tolerance of unrealistic experience emphasized in studies of cognitive control by Gardner et al., (1959). In the F test, found by the California group to be correlated with authoritarianism, one of the elements was intolerance of ambiguity. The F test was drawn on for those items in it which seemed most relevant to the specific concept of tolerance of ambiguity and ambivalence. To these were added a number of other items which seemed to sample aspects of behavior and experience relevant to tolerance of these and of complexity, unrealistic experience, and role confusion. A first form of this questionnaire, containing 20 items, was used in the preliminary research presented in the first chapter of this report. To increase the reliability of this promising measure, it was increased to 80 items. Of these, 8 items were in the end not scored, because they were also scored as a part of Barron's Independence of Judgment measure, and it was desirable to have the measures operationally independent of each other. The remaining 72 items were sufficiently diverse to suggest the possibility of being divided into several groups on the basis of content. They were divided into 5 groups, to see whether the sub-concepts represented could be separately measured; 4 items were omitted in this regrouping. The results are presented in Table 2. Tolerance of ambiguity, tolerance of ambivalence, and tolerance of unrealistic experience have intercorrelations in the same range as their reliabilities. They were therefore combined into a single variable which will be the next one to be discussed. The items on tolerance of role-confusion showed no evidence of consistency with each other and they were therefore not used. Tolerance of complexity showed internal consistency well above its correlation with the other variables and was kept as a separate variable.

The meaning of these variables must at present principally reside in the content of the items that make them up. To avoid lengthy presentation of entire questionnaires, a few items will be presented for each variable. They have been selected and will be presented in descending order of their correlation with total score on the variable. For most variables, measured by about 20 items, the six items showing the highest correlation with total score will be presented. For variables with a much larger number of items, a somewhat larger sample will be presented. The sign,+ or -, preceding the item indicates which response to the item was scored as indicative of the upper end of the variable. For tolerance of complexity the 6 items showing the highest correlation with total score (based on 20 items) are the following:

Table 2

Reliability (corrected split-half correlation) and Intercorrelations of Tolerance Measures

	Correlation with variable number 1	2	3	4	5	Number of items
1. Tolerance of ambiguity	.42	.22	.36	.18	.17	14
2. Tolerance of ambivalence		.25	.33	.01	.19	10
3. Tolerance of unrealistic experience			.46	.17	.19	12
4. Tolerance of role-confusion				-.06	.23	12
5. Tolerance of complexity					.55	20

+ 1. In so far as the study of philosophy makes one doubt his basic beliefs, it should be encouraged.

+ 2. Those religions are to be most respected which impose no uniform beliefs on their members.

- 3. The man who truly loves a woman must regard her as the best in the world in every important respect.

- 4. People fall very naturally into distinct classes, such as the strong and the weak.

- 5. Most of our social problems would be solved if we could somehow get rid of the immoral, crooked, and feeble-minded people.

+ 6. No one can be sure of conquering his difficulties; will-power is not enough.

Tolerance of complexity showed a correlation of + .23 with the easy esthetic judgment score and + .27 with the difficult esthetic judgment score. These correlations remain almost as high (+ .21 and + .25) when the art background variables are partialed out. Here is very clear evidence that esthetic judgment is positively related to tolerance of complexity as measured by self-report in a questionnaire, and that this relation is largely independent of the relation of each to background in art.

2. Tolerance of Ambiguity, Ambivalence, and Unrealistic Experience

Tolerance of ambiguity, ambivalence, and unrealistic experience was measured by 36 questionnaire items. The 10 items showing the highest correlation with total score are, in order:

+ 1. The expert ski jumper should enjoy the sport all the more if it remains a source of tension and even alarm.

+ 2. It would be exciting to arrive in a new city for the first time and find it enshrouded in a heavy fog.

+ 3. It is fun to get a new view of yourself--as with mirrors which show you your own profile, or in a frank account of other people's opinions of you.

- 4. Clouds that are frankly clouds and cover the whole sky are preferable to the little floating ones that leave you never knowing whether the next moment will be bright or dull.

- 5. Even if you were an expert in all the relevant languages, it would be disquieting to be in a multilingual group where you would never know which language to use next.

+ 6. Optical illusions and other experiences that put you in conflict about what is real and what isn't, are on the whole quite enjoyable.

- 7. To be wandering around a familiar city and suddenly realize that you are lost would be an unpleasant experience.

- 8. Anyone who lives from the income of investments would do well to settle for a smaller income rather than try for a larger income at slightly greater risk.

- 9. Regardless of whether it is necessary or not, it is preferable to have a clear schedule or plan for each day's activities.

- 10. It would not be very interesting to try to find one's way into what the experience of a blind person is like; it would be better not to think about it unless you had to.

The score derived from such items showed surprisingly low correlations with easy esthetic judgment, surprising in view of the results of the preliminary study. The correlation of +.13 was not statistically significant, nor was it significant when raised to +.15 by statistical control of art background. Difficult esthetic judgment shows greater evidence of a relationship, a correlation of +.22 which is increased to +.25 by statistical control of the art background variables.

3. Scanning

A third measure was of scanning, a dimension of cognitive control described by Gardner et al (1959). The upper end--high scanning--is defined as a tendency toward broad deployment of attention, so that one is acutely aware of what is going on outside of the main focus of attention and notices changes in background stimulation, unusual events of any kind, and also the possibly trivial elements in the events whose important aspects one is principally concentrating on. The lower end was thought of as narrow focusing of attention, so that little is noted outside the activity or event one is principally concentrating on, and even there little is noticed except the aspects most important or central to one's purposes or interests.

The variable of scanning seems likely to be related to an understanding of art, and hence to judgment, on two separate grounds. One is that understanding of a complex work of art involves the ability to attend simultaneously to various aspects of its stimulus value. The second is that interest in art (and hence an understanding of it) seems most likely to develop in scanners, since objects of art are constantly present as a part of one's environment from early childhood but may be reacted to solely for their practical significance except by the scanner. Chinaware and tableware, for example, are necessarily reacted to by everyone in eating, but who will notice their form and decoration? Magazines inform and influence one, but who will notice the arrangement of the type, and the color relationships to the advertisements? Young children early learn that the doorknob is an essential part of a house; who among them will first notice the shape of the doorway which frames it or of the space to which it leads? Perhaps the answer to each question is: the person who has a general alertness to a variety of aspects of the world around him, who notices things unimportant to the dominant interest of the moment as well as things central to it--in short, the scanner.

The internal consistency of the 20-item measure of scanning is reasonably satisfactory for the research use intended here, as the split-half coefficient is .54(corrected by the Spearman-Brown formula, as are the other split-half coefficients reported here). The items most highly correlated with the total score on scanning are the following:

+ 1. After having a date with a girl, I could describe pretty accurately the color and style of her clothes.

- 2. I am very sensitive to the emotional attitudes people sometimes want to convey but are unwilling to state openly.

- 3. When I am concentrating on one thing, I am really oblivious to everything.

+ 4. I seem to notice unusual noises sooner than other people do.

- 5. I rarely notice the color of people's eyes.

+ 6. Going down a street, I usually notice what the billboards and sighs say.

The expectation of a relationship here was confirmed. Scores on the self-rating questionnaire of scanning were for the entire 138 subjects correlated with the easy test of esthetic judgment to the extent of a coefficient of + .29. For those who took the difficult test of esthetic judgment, the correlation with it was +.28. Scanning is also correlated positively with the art background variables, so that when these are partialed out, the correlation with the easy test is reduced to the barely significant value of +.19, and the correlation with the difficult test almost disappears (+.07).

4. Sharpening

Sharpening is also a dimension of cognitive style developed in the work of Gardner and his collaborators. For present purposes a high degree of sharpening was defined as a keen awareness of differences, both at the time of occurrence and in later recall, among the various related or similar experiences occurring at different times--at the extreme, a sense of uniqueness about every experience. The lower end, referred to by some writers as levelling, is represented by a tendency toward awareness rather of similarities among various related or similar experiences occurring at different times--at the extreme, a great exaggeration of the similarities, and a fusing of memories, a blurring of the boundaries between them.

The split-half correlation of the sharpening score is .52. Of the 20 items included in the scale, those most highly correlated with total score were the following:

- 1. My memories of early childhood are of a general feeling rather than of detailed experiences.

+ 2. I can remember the name and looks of nearly every teacher I had in primary school.

+ 3. I have some fairly vivid memories of my first day in school.

+ 4. I have a clear mental picture of the passage of time since I was born, with major events of my life and world events pretty accurately placed.

- 5. I can hardly remember anything about popular songs I may have heard before the age of 10.

+ 6. Even without the evidence of photographs, I am directly aware of changes in my parents' appearance during my lifetime.

Sharpening might be related to esthetic judgment on the supposition that intense involvement in any work of art, and hence understanding of it, is favored by attention to the uniqueness of the experience being had while contemplating it.

This expectation was not appreciably confirmed. For the 138 subjects the correlation with the easy test of esthetic judgment was only +.06. For the subjects who took the difficult test, the correlation with it was +.16, suggesting a tendency toward a positive relation but not statistically significant.

5. Narrowness of Equivalence Range

Narrowness of equivalence range also derives from recent writings on cognitive style by Gardner and others. A tendency toward range of equivalence is defined as a tendency to react somewhat distinctively even to stimuli which are only slightly different. Broad equivalence range, the opposite end of the scale, is defined as a tendency to react to stimuli in an identical manner even when the difference between them is fairly considerable. There were difficulties in devising items for narrowness of equivalence range that seem clearly distinguishable from items on sharpening, since both involve elements of uniqueness vs. generality. An attempt was made to distinguish them by having narrowness of equivalence range not refer exclusively to temporal series of events and by having it refer more to the degree to which one acts on differences than to the degree of awareness of differences.

The split-half reliability coefficient for narrowness of equivalence range was only .38. The items with the highest correlation with total score were:

+ 1. Meaningless sounds may all be just noises, but it is very interesting to notice how much they all differ from one another.

+ 2. Words that are considered synonyms usually seem far from identical in meaning to me.

- 3. In studying psychology, I would be more interested in learning about the ways in which all people are alike than in learning about the ways in which people differ.

+ 4. Many more distinctions among colors are important to me then I have names for.

- 5. From knowing what groups a person belongs to, I feel I can get a pretty good idea of what kind of a person he really is.

+ 6. Each textbook I study seems distinctive in its approach.

The reason for expecting that narrowness of equivalence range might be related to esthetic judgment was essentially the same as for sharpening. In a different way this variable also refers to absorption in or enjoyment of the uniqueness of a specific experience as against reacting to it primarily in terms of generality or similarity to other experience. Such an approach might be expected to favor thorough understanding of a work of art.

The expectation too, was not appreciably confirmed. The correlation with the easy test of esthetic judgment was only +.145 and the correlation with the difficult test of esthetic judgment was nearly zero (+.03).

6. Flexibility

The notion of flexibility was also immediately suggested by recent writings on cognitive style. It is, however, a familiar concept, very similar to certain conceptions of intelligence, but applied specifically to cognition rather than to the ultimate adjustment which is influenced by it. A high degree of flexibility was defined as ability to change one's perceptual orientation when conditions make it appropriate, ability to shift the main focus of attention radically when doing so is adaptive, and toleration or liking of circumstances which call for exercise of this ability. Low flexibility was defined as the opposite; rigidity in the main focus of attention and set, and dislike or avoidance of circumstances which call for flexibility.

The internal consistency for the 20-item measure used here had the moderately satisfactory value of .52. The items most highly correlated with total score were the following:

+ 1. In a problem situation, I seem to be able to shift my attention around pretty fast from one aspect to another in search of a solution.

- 2. I get flustered when I have to attend to two different things at once.

- 3. I would rather stick to situations I already know how to cope with, and avoid situations where I know I will have difficulties.

+ 4. I find that giving attention to several different plans at once adds zest to my life.

+ 5. If I had made one set of friends in one context (in athletics, or some hobby or extracurricular activity) and an entirely different set of friends in another context, I would find it especially interesting to suddenly encounter a friend from each set simultaneously and get them both involved in a single conversation.

+ 6. In planning a lengthy paper, I like to consider several entirely different ways of organizing it before I finally decide on one.

The basis on which a correlation with esthetic judgment was expected for flexibility is that an esthetic mode of responding to objects is after all a somewhat distinctive way of considering objects which are required by the exigencies of everyday life, and that ability to take and enjoy an esthetic attitude would be favored by general flexibility of cognitive approach.

This expectation was also not confirmed. The correlation of this score with both the measures of esthetic judgment was slightly negative, although hardly different from zero: -.06 for the easy test of esthetic judgment and -.02 for the difficult test.

After the first 59 subjects had taken part in this research, a research assistant (David Singer) after familiarizing himself with recent writings on cognitive style went over the questionnaire items to try to judge what dimensions were intended to be represented by each item. He felt that there was insufficient discrimination in the case of some of the items for scanning, sharpening, and flexibility. As a result of this criticism, new items were made up for each of these dimensions and were included in the questionnaire given to the rest of the subjects, although not a part of the scoring of those dimensions as reported above. Separate scoring of these new items for these three dimensions yielded results which, on the whole, agree with those obtained with the original scales. For the new scanning items there is a definite tendency toward a positive relation, significant only for the easy esthetic judgment test; with it the correlation is +.21, while it is only +.13 for the difficult test. For the new flexibility items there is no acceptable evidence of a relationship, with a coefficient of +.11 for the easy test of esthetic judgment and +.08 for the difficult test. Similarly for the new sharpening items there was no evidence of a relationship, with a correlation of -.02 for the easy test and +.03 for the difficult test.

7. Field Independence

While the dimension of field independence was thought of at the beginning of the research as one possibly relevant to esthetic judgment, it was felt more promising to try to measure it by objective tests, since a multi-item objective test was readily available in the Gottschaldt Embedded Figures Test. Later it was felt that there might be some promise of obtaining a more general questionnaire measure, in view of the initial success obtained with questionnaire measures of other aspects of cognitive style. A 20-item questionnaire on field independence was therefore included in the questionnaires given to the last 79 subjects.

In devising items for this questionnaire field independence was defined as ability to maintain a set, to focus selectively on what one intends to do, and obtain new perceptual structuring, despite irrelevant attention-demanding stimulation and despite irrelevant well determined structure of the perceptual field or lack of structure. A high degree of field dependence, the opposite end of the scale, was defined by domination by strong or vivid stimulation and existing structure or by anxiety about lack of structure so that a set or intent contrary to these is overwhelmed.

This questionnaire measure of field-independence was internally consistent to about the same degree as the other 20-item measures of cognitive style, with a split-half coefficient of .53. The 6 items most highly correlated with total score are these:

+ 1. In a complex passage of orchestral music, with many instruments playing, it is easy for me to follow separately the theme being played by the solo instrument.

- 2. I sometimes have difficulty recognizing an acquaintance when I see him for the first time in a completely new setting.

+ 3. The fact that in steering a vessel you have to make your own path, instead of following an established road, would give sailing or motorboating a special appeal to me.

+ 4. Trying to find your way around and take a leading part in the group activity while blindfolded (as in some games) is a challenge I might well enjoy.

+ 5. If I were suddenly to fall into deep water in complete darkness, I am sure I would know immediately in which direction the surface was located.

+ 6. I am pretty good at sensing the concealed motives that lie behind a person's arguments.

The dimension of field independence had seemed relevant to esthetic judgment because of the supposition that the organization obtained in the experience of a work of art must be supplied partly by the individual's own discovery of form or structure not immediately and easily apparent, that the perception or understanding of this form often comes only as a result of careful, close attention directed to searching out such structure. This expectation was, however, not appreciably confirmed, as the correlation of this questionnaire measure with goodness of esthetic judgment was only +.13 for the easy test and +.06 for the difficult test.

As field independence is the last of the cognitive-style variables measured by questionnaires prepared especially for this research, this is an appropriate place to consider the relation of these variables to variables bearing the same names which might be measured by a battery of performance tests. The evidence is fragmentary because the study was not planned primarily to provide it. For some of the questionnaire variables, however, some of the performance tests described in the previous section of this chapter are among likely candidates for inclusion in a battery of such tests, and correlations between the questionnaire measures and these performance tests thus provide relevant though limited information as follows:

a. Tolerance of ambiguity, ambivalence, and unrealistic experience.

Three ratings of response to wearing aniseikonic lenses seem obviously relevant. The correlation of this questionnaire measure with effect obtained while looking at the hands was +.08; with effect on perception of the physical environment up to the point of looking at the hands was +.08; with maximum effect obtained in perception of the physical environment, +.16. The correlation with the sum of these ratings was +.16. While these correlations are not significant, all but one of the corresponding correlations with the items on unrealistic experience alone (+.04, +.18, +.18, +.22) do reach significance at the 5% level or better.

b. Scanning

A single performance measure obviously relevant to the concept of scanning is the number of words remembered incidentally in the Color-Naming and Color-Word Test. This measure showed no evidence of relation to the questionnaire measure of scanning ($r = -.01$, $N = 89$).

c. Narrowness of equivalence range

The only performance measure that seems possibly relevant to this concept is one of the response-set measures. It might be argued that with balanced questionnaires such as were used here a consistent tendency to agree or a consistent tendency to disagree with statements signifies an undiscriminating reaction to them as members of the class "questionnaire statements" rather than reaction to their specific content. If such an argument is used to justify the expectation of a negative correlation between agreement-disagreement response set and the questionnaire measure of narrowness of equivalence range, the expectation is disappointed, as the correlation has the insignificant value of -.07.

d. Flexibility

Three of the performance variables seem obvious candidates for inclusion in a battery to measure cognitive flexibility. Their correlations with the questionnaire measure of flexibility are all close to zero: conflict score on the Color-Naming and Color-Word Test, -.002; discoordination score on the Two-Hand-Coordination Test, -.006; non-constancy, +.074. However, none of these three measures have significant correlations with one another; this fact is reminiscent of the finding of Forster, Vinacke and Digman (1955), who also failed to find generality in potential measures of flexibility.

e. Field independence

The most clearly relevant measures here are those obtained from the Hidden Figures Test. The questionnaire measure of field independence was correlated +.25 with number right, -.20 with number wrong, +.04 with speed, and -.38 with the making of odd responses. These correlations are all in the direction that should be expected, and the first and fourth are statistically significant, the latter at the .001 level (N = 79). Since the questionnaire was given later than the Hidden Figures Test, these correlations might be thought to represent an influence of self-knowledge gained from this test upon questionnaire response. This interpretation seems unlikely, for speed is the one measure for which the subject clearly obtained self-knowledge, yet it is the one measure showing almost no correlation with the questionnaire variable.

Less persuasively, it might be argued that the questionnaire measure of field independence should be related to the Thurstone Hidden Digits score (r = +.10), to the other Thurstone tests of form perception (Dotted Outlines, r = +.13; Mutilated Words, r = +.13; Picture Perception, r = +.08), to the sum of all four of these scores (r = +.20), or to the Spatial Relations scores (number right, r = +.25; number wrong, -.08). While only the correlation of +.25 is statistically significant, all these coefficients are in the right direction.

For two of the five questionnaire scores for which any evidence is available, then, there is some definite indication of a tendency toward agreement with the performance measures. For the other three, the slight evidence available is not extensive enough to be regarded as strongly against such a tendency, but it offers no support for it either.

8. Independence of Judgment

Another important aspect of cognitive functioning is independence of judgment. For this purpose there was available a previously validated questionnaire measure. It consists of items found by Barron (1953) to be correlated with response to Asch's experiment on social pressure. In this well known experiment the subject is asked to make a fairly obvious judgment about a perceptual phenomenon, such as the relative length of two lines, when he is the last or nearly the last in a series of apparent subjects to give their judgment out loud, and the others are in fact confederates who by pre-arrangement give a judgment contrary to fact. Under this social pressure, many a subject will state a judgment contrary to the evidence of his senses. The scale for independence of judgment consists of those questionnaire items which Barron found to be most markedly correlated with the tendency to resist social pressure and stick to one's own judgment. Unlike the other questionnaire scales used here, this one does not contain an equal number of items phrased in the two directions; it contains 5 statements for which a positive response contributes to the independence score, and 11 statements for which a negative statement contributes to it. The alpha coefficient of internal consistency for this 16-item scale, for the 138 subjects used here, was .51. The items showing the highest correlation with total score are the following:

- 1. What the youth needs most is strict discipline, rugged determination, and the will to work and fight for family and country.

- 2. The happy person tends always to be poised, courteous, outgoing, and emotionally controlled.

- 3. If young people get rebellious ideas, then as they grow up they ought to get over them and settle down.

+ 4. Some of my friends think that my ideas are impractical, if not a bit wild.

- 5. I must admit that I would find it hard to have for a close friend a person whose manners or appearance made him somewhat repulsive, no matter how brilliant or kind he might be.

+ 6. The unfinished and the imperfect often have greater appeal for me than the completed and polished.

This is a particularly interesting aspect of cognitive function in relation to esthetic judgment. On the one hand, if one supposed goodness of esthetic judgment as measured in this and other research to be a function primarily of the extent to which one has sought out and accepted the judgment of experts on esthetic values, then one might expect that goodness of esthetic judgment would be negatively correlated with independence of judgment generally. If one supposed, on the other hand, that goodness of esthetic judgment derives to a greater extent from having observed works of art carefully and formulated one's own judgment about them, so that the careful observer of works of art tends to arrive independently at judgments of value similar to those which experts have arrived at largely by the same process, then one might expect a positive correlation between goodness of esthetic judgment and general independence of judgment. An hypothesis underlying this research was that goodness of esthetic judgment would be in considerable part an outcome of independent discovery of the same facts about works of art that the experts have made and that a positive correlation was likely.

This expectation was rather fully confirmed. The correlation with the easy test of esthetic judgment for the entire sample of subjects was +.25, and the correlation with the difficult test of esthetic judgment for the 69 subjects who had taken it was +.33. These findings are almost entirely independent of background in art; the partial correlations, with background variables controlled, are +.225 and +.360, one slightly smaller and one slightly larger than the corresponding simple correlation.

The findings on independence of judgment confirm, in a sense, the tenor of findings reported already for tolerance of complexity and for tolerance of ambiguity, etc., for the measures are interrelated. Independence of judgment is correlated +.35 with tolerance of complexity and +.44 with tolerance of ambiguity, etc.

9. Regression in the Service of the Ego

Another dimension of cognitive functioning which was explored through questionnaires was regression in the service of the ego, through a questionnaire devised by David Singer in connection with unpublished research of his own. (A few items in his questionnaire which referred to the arts were for present use replaced by new items not referring to the arts.) This questionnaire became available only for the last 49 subjects of this investigation.

This concept is derived from recent psychoanalytic writings on ego psychology and its application to the psychodynamic meaning of arts. The concept has been especially applied in discussions of creativity, but seems applicable also to appreciation of the arts. The idea as presented by Kris (1952), Schafer (1958), and others may be summarized by saying that creation or enjoyment of the arts involves in part a regression from more mature forms of cognitive activity to less mature forms, but a controlled regression in which the ego functions remain intact and in command, so that a positive use of the less mature functions can be made in the interest of more mature functions. It seems likely that just such an ability to fuse the more infantile with the more mature might be involved in the ability to enter fully into a work of art and to integrate it into one's general cognitive functioning.

Like the independence of judgment scale, the regression measure was not perfectly balanced; it included 14 positive items and 11 negative items. Its split-half reliability was .52. The 6 items most highly correlated with total score were:

+ 1. Unusual but unimportant aspects of a situation or object often intrigue me, occupying my attention and imagination for a time.

- 2. I get little pleasure or fun out of playing with words or language--as by talking nonsense, baby-talk or in a foreign accent--and seldom do that sort of thing.

+ 3. I sometimes have daydreams in which I become a (heroic' type of figure; either all-powerful, all-knowing and successful or someone who has sunk to the lowest depths of depravity, weakness and suffering.

+ 4. I find it easy when playing with very young children to get into their world and try to experience things as they do.

+ 5. I sometimes just let my thoughts wander aimlessly, and find myself daydreaming and thinking about all sorts of unusual and unrelated things.

- 6. I can detect in myself no strong anti-social impulses of the sort which, under certain circumstances, might lead to crime.

So far as the evidence permits with the restricted number of subjects involved (N = 40), the expectation of a correlation with esthetic judgment was very amply fulfilled. The correlation of this dimension with the easy test of esthetic judgment was + .29 and the correlation with the difficult test of esthetic judgment was +.33. These relationships, especially the first, are sufficiently associated with a relationship to art background so that when this is controlled the partial correlations fall to +.20 and +.30; only the latter value is still significant at the 5% level.

10. Jungian Measures of Cognitive Orientation

Jung, in his Psychological Types (1923), suggested dimensions of cognitive functioning somewhat akin to some of those discussed recently in connection with the concept of cognitive style or cognitive control. The dimension of introversion vs. extroversion is the one which has received most attention. It was primarily in order to attain an especially promising measure of that dimension that, in the preliminary investigation reported in the first chapter, the Myers-Briggs Type Index (Myers, 1962) was administered. While the expectation, based on previous research, that introverts might be especially likely to show involvement in art was not confirmed, two of the other Jungian dimensions measured by the questionnaire did turn out to be appreciably related to the esthetic quality of personal preferences in painting. The same questionnaire was therefore administered to the subjects in the present investigation.

In this study, in contrast to the preliminary one, there was some evidence of a relation to introversion. The easy test of esthetic judgment showed a correlation of -.22 with the dimension of extroversion vs. introversion, and the partial correlation obtained by controlling art background variables rose to -.29. This highly significant relationship was not confirmed with the difficult test of esthetic judgment, where the correlation was a negligible one in the same direction (-.04).

As in the preliminary investigation, the dimension of intuition vs. sensation showed a marked correlation with esthetic interests, a coefficient of +.36 in the easy test of esthetic judgment and a correlation of +.24 with the difficult test of esthetic judgment. These correlations are both associated with the relation of the variables to art background; the partial correlations with art background controlled are only +.21 and +.11, the former still statistically significant.

The previous finding of a relation with the dimension of perception vs. judgment was confirmed to some extent for the easy test, with a coefficient of +.22, but was not confirmed for the difficult test, with the negligible correlation of +.09; these correlations also drop somewhat, to .18 and .07, when art background variables are statistically controlled. The findings of this research resemble those of the previous study in finding no evidence of a relation to the dimension of feeling vs. thinking, the correlation being +.08 for the easy test and +.09 for the difficult test.

The suggestion of a connection between goodness of esthetic judgment and a tendency toward intuition and toward perception, as these are defined by the authors of the test, seems quite reasonable. The dimension of intuition vs. sensation is defined in part as involving an orientation toward understanding an object in full for its own sake, as against willingness to tolerate purely external or sensory apprehension of it. The same might be said for the rather paradoxical correlation of goodness of esthetic judgment with a tendency toward perception rather than judgment. The perception end is defined as an interest in unique understanding of the objects, whereas the judgment end is defined as an interest in quickly making a judgment, even if necessarily a superficial one, about whatever one is in contact with. Thus, the judgment end corresponds partly in defined meaning to intolerance of ambiguity, ambivalence, or complexity. It also seems to correspond in part to the idea of dependence of judgment, since the sort of judgments referred to often are those taken over already formulated by others.

These findings are in fact not independent of those on the other dimensions of cognitive style. Both intuition and perception as measured by the Myers-Briggs test are correlated positively with independence of judgment (.38 and .30 respectively), with tolerance of complexity (.26 and .62) and with tolerance of ambiguity, ambivalence and unrealistic experience (.06 and .34).

F. Other Variables of Personality

1. Extroversion versus Introversion

A separate set of items in the questionnaires constructed especially for the present research program was concerned with extroversion vs. introversion. This set was used in the preliminary research reported in the first chapter. It was retained despite use of the Myers-Briggs Type Indicator, which includes a measure of extroversion vs. introversion. The Myers-Briggs scale, as described by its authors (Myers, 1962), appeared to be directed at the intellectual or cognitive aspects of extroversion vs. introversion, whereas the items in the author's questionnaire stress social aspects. The items were derived from previous questionnaires intended to measure this dimension, most directly from the Maudsley Personality Inventory (Eysenck, 1956), but with deliberate balancing of the items so that an equal number would be in each direction.

The following 6 items showed the highest correlation with total score on the 20-item questionnaire:

- 1. I am inclined to keep in the background on social occasions.

- 2. The presence of the opposite sex makes me feel shy.

+ 3. Other people probably regard me as a lively person.

+ 4. I am rather a talkative person.

- 5. I am inclined to be rather quiet when in a large group.

- 6. I could be quite happy even if circumstances prevented me from having many social contacts.

With extroversion defined this way, the expectation of a negative correlation between it and goodness of esthetic judgment is perhaps based on a popular stereotype of the socially-isolated pursuit of beauty in an ivory tower, rather than on any psychological theory. In any event, the expectation is definitely not confirmed. The correlations were very small and were in fact in the opposite direction, a tendency toward extroversion being correlated +.13 with the easy test of esthetic judgment and +.07 with the difficult test.

2. Anxiety

Anxiety, or essentially what in some questionnaire usage has been called neuroticism, was measured by 20 items which had also been employed in the preliminary research reported in the first chapter. Like the items on extroversion, these were drawn largely from previous questionnaires, though more modification and invention were required here to obtain an equal number of items stated in each direction. The 6 items most highly correlated with the 20-item total were the following:

+ 1. I often experience periods of loneliness.

- 2. I spend very little time worrying over possible misfortunes.

- 3. My mood is pretty steady, not characterized by frequent ups and downs.

+ 4. My feelings are rather easily hurt.

+ 5. I often find that I have made up my mind too late.

+ 6. There are quite a few subjects on which I am rather touchy.

The variable of anxiety was included in the research without any definite expectation about a relation to esthetic judgment, primarily to serve some incidental purposes in exploring personal preferences and expressive behavior. Like the other variables to be mentioned later which were included for this reason, it was also looked at in relation to esthetic judgment, and slight evidence of a positive relationship was found. The correlation with the easy test of esthetic judgment was +.18 and with the difficult test of esthetic judgment +.16. Statistical control of art background variables left these correlations almost exactly the same (+.18 and +.14), and only the larger of the two approaches significance at the 5% level. Such slight evidence as this provides of a relation between anxiety and goodness of esthetic judgment the author would be inclined to interpret tentatively in relation to the finding of regression in the service of the ego. That is, the person with greater interest in esthetic values may be one who is more able to face and recognize unpleasant emotions and undesirable states in himself, and he may express this tolerance in his answers to questionnaires as well as in his willing exploration of the unpleasant as well as the pleasant in the world of art.

3. Dimensions of Temperament

In the beginning of this program of research it seemed likely that the dimension of temperament described by Sheldon (1942) as cerebrotonia might be relevant in the same way that intellectual introversion seemed to be. In fact this variable as described by Sheldon has a good deal in common with the concept of intellectual introversion. In some earlier research by Child (1950), Sheldon's concepts were drawn on in formulating a self-rating questionnaire on this and the two other dimensions of temperament which Sheldon has described, viscerotonia and somatotonia. The items which in that research were found to be somewhat correlated with dimensions of physique, in accordance with Sheldon's claims, clearly did not form three independent dimensions, and there was a tendency for cerebrotonia to be represented principally by items which contrasted either with viscerotonia or with somatotonia. In the present research, therefore, an attempt was made only to measure viscerotonia and somatotonia, with the lower end of each representing some of the characteristics Sheldon groups together under the label of cerebrotonia.

In the preliminary research reported in Chapter 1, viscerotonia was found to be negatively correlated with esthetic preference--that is with a tendency for one's preferences among paintings to be related positively to judgments of esthetic quality of the same paintings by experts. Somatotonia was not found to be related in either direction.

In the present research the results for viscerotonia are in the same direction as before, though not of large magnitude. The easy test of esthetic judgment shows a correlation of -.19 and the difficult test a correlation of -.22 with viscerotonia; when art background is partialed out, the former coefficient rises to -.21 while the latter drops to -.17. For somatotonia there is this time a very slight indication of a relationship in the predicted direction with the easy test, with a correlation of -.15 but this is diminished to -.07 by taking art background into account, and the correlation with the difficult test is exactly zero. These findings, when put together with the similar findings in the previous study, indicate that among college students an interest in esthetic values is negatively related in small degree to the somewhat passive sensual enjoyment of relaxation and ease that is a part of the concept of viscerotonia, but shows no sign of incompatibility with the vigorous, active kind of interest that is represented by the concept of somatotonia.

4. Overt Concern for Performance

The experimenter who conducted the individual sessions with the last 79 of the subjects was impressed by the great variations among the subjects in their apparent concern about how well they did in the various perceptual tests that were conducted in the individual sessions. She accordingly made ratings of this overt concern, and these ratings were correlated with the measures of esthetic judgment. The result was a correlation of -.21 with the easy test of esthetic judgment and a correlation of -.26 with the difficult test of esthetic judgment. This finding may be viewed as consistent with the results obtained with the questionnaires on cognitive functioning. Overt concern with how well one did on the test seems likely to be related to great concern with the evaluation of oneself by the other persons with whom one is in contact, in contrast with primary concern for one's own self-evaluation. This finding may then reflect the relation of esthetic judgment to general cognitive independence. Against this view is the fact that the rating of concern had almost a zero correlation (+.06) with the questionnaire measure of independence of judgment.

5. Scholastic Aptitudes

In the preliminary research on esthetic preference, tendency for one's personal preferences to agree with esthetic evaluation by experts was somewhat related positively to verbal aptitude as measured by the C.E.E.B. Scholastic Aptitude Test, but not related to mathematical aptitude. Evidence in the same direction but much less pronounced was obtained in the present study. Most of the subjects had already taken, at the time of application for college, the two aptitude tests of the College Entrance Board; from calculations based on these scores, we omitted several subjects for whom English appeared not to be the native language. The verbal score was correlated +.17 with the easy test of esthetic judgment (N = 129), and only +.10 with the difficult test of esthetic judgment

(N = 71); the former coefficient rises to +.24 when art background is controlled, while the latter coefficient drops slightly. Both correlations were almost zero for the mathematical aptitude score, -.01 for the easy test and -.03 for the difficult test.

G. Measures of Preference and Expressive Style

In conjunction with this investigation, while measures were being collected on a number of subjects, the author wished to explore several hypotheses about stimulus preferences and characteristics of expressive production, hypotheses having nothing to do with esthetic judgment. Since measures of several variables were being obtained for these other purposes, we studied their relation to esthetic judgment. Since some of the same tests were used, we will also report here an unsuccessful attempt to measure originality, although an equally appropriate place would be with the performance measures of cognitive style. Despite the purely empirical character of most of this enquiry, several significant relationships were found. The first group of measures to be discussed here have to do with masculinity and femininity.

1. Masculinity-Femininity

The dimension of masculinity vs. femininity seems eminently worth exploring in any study of behavior in relation to the arts. This is because of the seemingly contradictory ways in which sex differences appear to enter into such behavior. On the one hand, artistic creativity has been and continues to be displayed primarily by men. Artistic appreciation and criticism, too, have been very largely masculine pursuits. On the other hand, there is a common belief in the United States that interest in or participation in the arts is essentially feminine and somewhat threatening to the male sense of self-esteem. More specifically, much that has been written about psychoanalytic interpretation of the arts and much in the way of obvious facts about the social context of some of the arts at present indicates some sort of tendency toward bisexuality or at least greater freedom of sexual role in some people actively participating in them.

To this issue of such great interest the findings of our research, unfortunately, seem to make nothing but a very inconclusive negative contribution because of the nature of the measuring instruments used.

One measure was a questionnaire comprising items which tend to discriminate between the sexes. The items were drawn from those of Gough's Masculinity-Femininity Test (Gough, 1952). All the items of Gough's test were embedded in one of the personality questionnaires already mentioned. Since they included an excess of items for which a positive response was scored as feminine, enough were eliminated in scoring so as to produce a balanced set--28 items, 14 scored in each direction. This questionnaire has been used as a measure of more or less conscious sexual identification (Miller, Swanson, et al., 1960). It does not in the present research, however, promise to be a good measure of anything, as its internal consistency is very low; a split-half correlation, corrected, has only the value of +.25. Whether this low reliability is typical or results from characteristics of our subjects or from the items' being embedded in a different setting and answered on a 4-point scale (collapsed for scoring into a dichotomy) is not known. At any rate, as might be expected, it shows little relation to esthetic judgment. Its correlation with the easy test is +.02, and with the difficult test +.15. (This is in the direction of good esthetic judgment's going with femininity, but does not even approach significance. All the measures reported in this section have been scored with femininity as the upper end of the scale.)

A second measure of masculinity vs. femininity was provided by the Franck Drawing-Completion Test, which has been used as a measure of unconscious or latent sexual identification (Franck & Rosen, 1949; Miller, Swenson, et al., 1960). Here the subject is presented with 36 incomplete drawings, each to be quickly completed in whatever way he pleases. The test was introduced to the subjects of the present research as a test of originality (not a standard mode of introducing it). It was then scored by the manual prepared by Franck (n.d.), based on classifying a person's completion of an item as more closely resembling responses typical of man or responses typical of women. Inter-judge agreement in using the manual was very satisfactory (uncorrected r of +.78 for the first 59 subjects, +.70 for the other 79 subjects). One judge rated one group of subjects at one time, the other group a year later, and his absolute standards shifted. We therefore used only the ratings of the other judge, who made them all at one time. The internal consistency of this test as scored by this judge was only moderately satisfactory; the corrected split-half reliability coefficient is .43. The measure shows no evidence of a relation to esthetic judgment, with an identical correlation of -.04 with the easy test and the difficult test (in the direction of masculinity's going with good esthetic judgment).

Two other measures were ratings of the masculinity-femininity of visual material produced by the subjects--that is, of the degree to which their drawings were in themselves masculine or feminine in connotative meaning. First, this approach was applied to the drawing completions the subjects produced on the Franck test. The individual items were rated on a 3-point scale, and a person's score was obtained by adding his item scores. Inter-judge agreement for the subjects' scores was extremely poor; inter-judge correlations were .11, .05, and .04, yielding a corrected reliability coefficient of only .18 for the pooled rating. The internal consistency of the subjects' scores as obtained from the pooled ratings of the 3 judges, on the other hand, was very good, with a corrected split-half coefficient of .77. This measure is nearly independent of the two already described, with a correlation of +.11 with the Gough score and +.15 with the Franck score. (These two had a correlation of +.13 with each other.) It resembles these in showing no significant relation to either the easy test (r = -.10) or the difficult test (r = -.12) of esthetic judgment; both are in the direction of masculinity of production going with good esthetic judgment.

Finally, this same rating procedure was applied to a single abstract drawing the subject was required to produce during 5 minutes of a group session, as a task intended to measure originality. There was some agreement among judges (inter-judge correlations of .70, .42, and .34), but subject consistency could, of course, not be assessed. This measure had no significant correlation with the questionnaire measure (+.11), a nearly significant correlation with the rated masculinity-femininity of drawing completions (+.19), and a barely significant correlation of +.22 with the score of tendency to resemble women in drawing completions made. Thus it shows a little more sign that do any of the other 3 measures of representing something in common among them. But it too gives no evidence of being related to esthetic judgment (r is -.07 for the easy test and +.04 for the difficult test).

We must conclude that our study provides no evidence that esthetic judgment is related in either direction to a dimension of masculinity-femininity. For lack of sufficient knowledge about the meaning of the measures used, it is impossible to judge whether this is satisfactory negative evidence.

2. Originality

Ratings of originality were made on the subjects' responses to three tests which have already been described: the Barron Inkblot Test, the Franck Drawing-Completion Test, and the 5-minute Abstract Drawing. In addition a single 5-minute application was made of the "New Uses of Common Things" test (see Guilford et al., 1954); the subjects were required to write down as many and as varied uses as they could for "a book", and these responses also were rated for originality. Inter-judge agreement was not very good. Subjects' scores for originality on the Inkblot Test based on different judges' ratings gave a mean correlation of only +.37; on the Drawing-Completion Test, two judges intercorrelated +.50; on the Abstract Drawing, the mean intercorrelation among judges and +.35. Only on New Uses for Common Objects was there good agreement, with two judges intercorrelating +.75. The intercorrelations among these 4 possible measures of originality were low and insignificant. Two of the six coefficients were negative, and the highest was +.15.

The principal conclusion we can draw is that no general dimension of originality has been satisfactorily measured by these tests. The correlations with the esthetic judgment tests and also such as to suggest that what is being measured may be something specific to artistic or semi-artistic responses or stimuli. Neither Inkblots nor New Uses for Common Things show any but a near-zero relation between originality ratings and the esthetic judgment tests. Drawing-Completions originality, on the other hand, yields significant correlations of +.19 with the easy test and +.25 with the difficult test. Abstract-Drawing originality agrees in direction (+.07 and +.17), although the correlations are not significant. The abstract drawings by the subjects were also rated for their esthetic merit and for their technical skill. These ratings gave correlations of +.30 and +.23 with the difficult test, and smaller correlations of +.16 and +.05, neither significant, with the easy test. These additional facts suggest that at least for the abstract drawings the "originality" measure may tend toward relation with esthetic judgment for reasons other than its possible measurement of originality, since the "originality" rating of the abstract drawing has substantial correlations of +.36 and +.45 with the ratings of esthetic merit and of technical skill. (Comparable ratings of the drawing completions have not yet been finished; the originality ratings of the drawing completions are not, however, correlated with the esthetic and technical ratings of the abstract drawing.)

3. Color Preferences

During the individual session all 138 subjects were given a color-preference test based on that devised by Barrett & Eaton (1947) but considerably modified. The subject was presented with pairs of small patches of colored paper, each pair mounted on a black 3" x 5" card. The two patches were always approximately of the same hue but differed in one of three ways: One was a "Hue" and the other a "Tint" of it; or one was a "Hue" and the other a "Shade" of it; or one was a "Tint" and the other a "Shade." There were 24 pairs to represent each of the 3 kinds of choice, presented in mixed order. The subject was not told how the colors in a pair differed and was asked only to express his immediate preference within each pair. Each subject was later given the Ishihara Test, and the records of 8 who showed some sign of color deficiency were deleted in treating the results from the color-preference test (and also other measures which were specifically concerned with color). Scores on this test showed remarkably high correlations with the esthetic judgment scores. Preference for hues over tints had a correlation of only +.17 with the easy test and +.04 with the difficult test. But preference for hues over shades had correlations of -.15 and -.28, while preference for tints over shades had correlations of -.32 and -.41. These last 3 correlations are all statistically significant, the last two highly so (N is 130 for the easy test, 76 for the difficult test). The relationships indicated are only slightly associated with a joint correlation with art background. The partial correlations with preference for tints over shades, art background controlled, retain the highly significant values of -.29 and -.35 for the easy and difficult tests respectively.

What these findings indicate is a tendency for persons with higher esthetic judgment scores to choose shades over tints and to some extent shades over hues more frequently than do persons with lower esthetic judgment scores. This finding of some relationship between simple sensory preferences and esthetic judgment seems very surprising. But it is of great importance if it is generally valid.

In an effort to begin clarification of this finding, an additional color-preference test (which will be fully described in a later publication) was devised for use with the last group of subjects. This additional test, instead of presenting contrasts within the framework of artists' classifications of colors, presented contrasts of clearly defined nature in relation to the psychological dimensions of hue, saturation, and brightness. Contrasts were presented for choice along each of these dimensions, separately on a black background and on a white. Only for preferences along the brightness dimension was there an indication of a relation to esthetic judgment. Preference for the color of lower brightness, of two patches on a black background, had an insignificant correlation of +.13 with the easy esthetic-judgment score; but for colors on a white background the corresponding correlation had the significant value of +.30 (N = 47). When art background variables are partialed out, this correlation falls to +.276, not quite at the 5% significance level. The correlations with the difficult esthetic-judgment score were close to zero (+.03 and +.08). These results in themselves are not impressive. There remains, however, some reason to believe the relationship suggested is genuine. An additional 32 subjects (not included in the 138 of the main study) were given this new color-preference test and the easy esthetic-judgment test. Their tendency to prefer the color of lower brightness was correlated with the esthetic judgment score +.20 when a black background was used and +.35 with a white background. Only the larger of these is almost significant at the 5% level, but they agree in direction with the previous correlations. They also agree with the statistically more decisive finding of a preference for shades with the test applied to all subjects in the main study. The largest and most consistent way that shades differ from hues and tints is in being of lower brightness (i.e., value, in the terminology of the Munsell color system). This difference is especially pronounced when a shade is paired with a tint, and it was for that comparison that the most decisive evidence of a relation to esthetic judgment was found.

For subjects such as were used in this study, then, esthetic judgment is positively correlated with a tendency to prefer colors of lower brightness when paired colors differing in brightness are offered for choice. The speculations that might be based on this finding will be of greatest interest when they can be combined with further pertinent observations.

Other Visual Preferences

With the first 59 subjects, a test of visual preferences was used as an additional measure possibly relevant to masculinity-femininity. The stimuli consisted of 20 of the same pairs of abstract paintings used for a measure, described earlier, of perception of sexual connotative meaning. They were pairs of abstract paintings so selected that, in the opinion of the selections, one would be definitely masculine in connotative meaning and the other definitely feminine. They were presented without mention of this characteristic, and embedded in 40 other pairs with less decided sexual connotation, and the subject was asked to record his personal preference for one painting in each pair. A count was made of the number of pairs on which the subject preferred the more masculine painting. As this score showed no tendency to be related to either of the measures of masculinity-femininity adapted from previous research (the Gough and Franck tests), it was one of the measures chosen for subsequent elimination to make way for inclusion of the Difficult Esthetic Judgment Test. The choice was unfortunate, for this preference measure yielded some other results which are of interest and which can not be confirmed until the measure is used in further investigation. Most important for present purposes is the fact that in these 59 subjects a preference for the more masculine painting was correlated +.29 with the easy esthetic judgment score. Such a relationship may have something to do with the general sexual element in the choices made. Or it may have to do much more specifically with strength, if people of good esthetic judgment are likely to prefer strength in painting.

Another measure based on visual preferences is the Barron-Welsh Art Scale (Barron & Welsh, 1952), which was administered in the individual session. The subject is given a pile of cards on each of which is printed a different abstract drawing. He is asked to divide them into those he likes and dislikes (and in our administration, is asked to adopt a cutting point which will make the division approximately equal). The drawings are of two kinds: those liked more by artists than non-artists, and those liked more by non-artists than by artists. The score represents the subject's tendency to make the choices more characteristic of artists. This measure has been found to be positively correlated with indices of originality (Barron, 1958). It might thus have been reported on earlier in the section on originality. The Barron-Welsh Scale is not in our data appreciably correlated with the measures of originality which were described in that section because of their seemingly obvious relevance to originality (the correlations vary from -.01 to +.10). It is therefore reported on here; quite possibly, however, it is a much better index of a general dimension of originality than are those other measures, which were found to be hardly appreciably correlated with each other.

The Barron-Welsh Art Scale is significantly correlated with the easy esthetic judgment test ($r = +.18$, $p < .05$). This correlation may well be connected entirely with a common relation to variables of art background, for when these are controlled it drops to +.09. The correlation of the Barron-Welsh Art Scale with the difficult test is in the same direction but not significant ($r = +.11$).

Finally, as a beginning of an inquiry into the contemporary meaning of personal preferences for one historical style over another, a measure was obtained of relative preference for Baroque vs. Classical Art. These terms are used here as defined by Wölfflin (1932), with "Classical" meaning "High Renaissance," not ancient. Pairs of contrasting works were presented for choice with no statement about the nature of the contrast; these pairs were based on the illustrative examples given by Wölfflin. A surprisingly high degree of personal consistency was found; the corrected split-half correlation was +.62. The measure of preference for Baroque over Classical showed an entirely unexpected correlation of +.22 with the easy esthetic judgment test and an insignificant correlation of +.14 with the difficult test. When art background is partialed out, both of these correlations have the value +.14. The unexpected positive finding here has several possible interpretations, and a choice among them would depend on further study. One is that the Baroque member of most pairs presented tends to be the one which our experts would consider esthetically better (whether the reasons lie in Wölfflin's biased selection of examples or in style prejudices of our experts). Another is that the Baroque works are more complex or else less like present-day art, and therefore require, in comparison with Classical, more background or interest in art in order to be understood and thus potentially liked.

SUMMARY AND DISCUSSION

We have reported a method of measuring esthetic judgment, or the extent of agreement with expert opinion on the relative esthetic value of the two works of art presented in each of a number of slide pairs. Measuring esthetic judgment in this way in 138 male university students working as paid subjects in psychological research, we have reported the relation of their esthetic judgment to a number of other variables.

First, we find that esthetic judgment is related to amount of background in art. We have no way of assessing to what extent this reflects an influence of background on esthetic judgment score and to what extent it reflects a tendency for people with better esthetic judgment to be more interested in art and thus obtain more experience with it. Since background in art might conceivably be an influence both on esthetic judgment and on the other variables to which we relate esthetic judgment in the rest of this chapter, we have in every case determined the partial correlation between variables, with art background statistically controlled. Many of the relationships we report remain statistically significant, and some become even more so, when art background is thus removed as a possible joint cause; some are greatly diminished. It would be premature to draw any conclusions from this one study about which relationships exist only by virtue of a common relation with art background and which exist independently of variations in art background. It is perfectly clear, however, that the relationships we report, taken as a whole, cannot be attributed to a common origin in variations in art background.

Second, for several groups of measures we fail to find consistently the significant relations to esthetic judgment that various hypotheses might reasonably have led one to expect. These are as follows:

(a) Skill in perceiving visual form. Here we find a predominance of correlations in the predicted direction, but none are large and few are even barely significant statistically.

(b) Skill in perceiving human meaning in ambiguous stimuli. Our results with different measures are contradictory in direction and none are statistically significant.

(c) Masculinity vs. femininity. Several measures used showed no sign of relation to esthetic judgment; on the other hand, their value as measures of masculinity-femininity is doubtful.

(d) Originality. Of several behavioral measures of originality, the only ones showing a relation to esthetic judgment were ones having to do with the subject's own graphic productions; thus the suggestion is of a factor specific to art rather than of general originality.

(e) Behavioral measures of cognitive style, other than ones classified under the more specific headings above. Here the correlations of several measures were small and sometimes in the opposite direction from what was predicted. There was one outstanding exception where results were so striking and consistent as to suggest strongly a genuine and substantial relationship to esthetic judgment: reaction to putting on aniseikonic lenses. This result indicates a positive relation of esthetic judgment to a behavioral measure of tolerance for unrealistic experience.

For none of these groups of measures is it true that the measures used are a well-established set known to give reliable and valid coverage of a group of intercorrelated variables. Thus we are unable to view the results as providing strong evidence that esthetic judgment is not importantly related to the aspects of behavior referred to. All we conclude is that our evidence is against any important relation of esthetic judgment to these aspects of behavior as we have been able to measure them.

Third, for a number of aspects of cognitive style as measured by questionnaires, significant relations with esthetic judgment were found. Esthetic judgment showed tendencies to be positively related to the following questionnaire measures:

(a) Tolerance of complexity (accompanied by non-significant correlations in the same direction for tolerance of ambiguity, ambivalence, and unrealistic experience)

(b) Scanning

(c) Independence of judgment

(d) Regression in the service of the ego

(e) Intuition rather than sensation

(f) Perception rather than judgment

The positive ends of these variables, taken together, suggest a person of actively inquiring mind, seeking out experience that may be challenging because of complexity or novelty, ever alert to the potential experiences offered by stimuli not already in the focus of attention, interested in understanding each experience thoroughly and for its own sake rather than contemplating it superficially and promptly filing it away in a category, and able to do all this with respect to the world inside himself as well as the world outside. The consistency of this picture is marred somewhat by the lack of significant findings with some additional questionnaire measures which would be consistent with this general picture. However, the results obtained are very impressive. They are in general agreement with the results of the more limited preliminary investigation reported in the first chapter, which used different subjects and a measure of esthetic preference rather than esthetic judgment. They also are strikingly similar in some respects to the picture of personality correlates of creativity which has emerged from the Berkeley studies of more vs. less creative people in several professions (Barron, 1958; MacKinnon, 1962).

Fourth, a relation to esthetic judgment was found for several measures which one can not fit so confidently into the pattern just described. Some perhaps do properly form a part of it, and some may be entirely separate correlates of esthetic judgment.

(a) Anxiety, as measured by a questionnaire, shows a positive relation to esthetic judgment. For persons high in esthetic judgment, a high anxiety score may reflect not an unusually high level of anxiety but an unusually high willingness to be aware of anxiety and report it. (This interpretation is consistent with the substantial correlation of +.50 between Anxiety and Regression in the Service of the Ego for the 49 subjects for whom the latter measure is available.)

(b) A negative correlation between esthetic judgment and viscerotonia (love of comfort and relaxation) is perhaps consistent at a temperamental level with the picture of an actively inquiring mind interested in challenges. Such an interpretation might equally call for a positive relation with somatotonia, which was not found. At any rate, the result with viscerotonia agrees with results on esthetic preference and viscerotonia obtained in the preliminary study of Chapter 1.

(c) A negative relation between esthetic judgment and overt concern for performance in the individual test session may fit into the general pattern if it reflects, for those high in esthetic judgment, relatively greater concern for their own self-evaluations and relatively less concern for the impression they make on other people.

(d) A small but significant positive correlation between esthetic judgment and verbal aptitude agrees with the findings of the preliminary study but seems to have no distinctive bearing on the general interpretation offered here.

(e) Esthetic judgment was found to be positively related to several measures of visual preferences: preference for colors of lower brightness ("shades"), preference (on the Barron-Welsh test) for abstract designs liked by artists, and preference for Baroque over Classical art. What the person of better esthetic judgment prefers in each case may possibly be said to have in common a relative lack of obvious unity, structure, or meaning. This seems clear in the case of Baroque style compared with Classical. It also seems clear in the case of the Barron-Welsh test where the designs preferred by artists are asymmetrical and chaotic in comparison with the generally simple geometrical figures preferred by non-artists. The results on color preferences may have a similar meaning if, as Knapp says in a different context (1958), more somber colors have less of a tendency to impel attention and a specific meaning-response and are thus more challenging to the viewer.

This study has dealt with individual differences within only one particular group: men who are students at a particular university. Whether the results can be extended to other groups cannot be known except when other groups are studied in some similar way. The results of this study constitute tentative findings which might well serve as sources of hypotheses in the study of other groups. Of special importance is the general pattern into which a number of findings fall, suggesting that good esthetic judgment is in important part an outcome of a general cognitive approach to the world, an approach involving search for complex and novel experience which is then understood and evaluated through relatively autonomous interaction of the individual with objects which provide such experience. The esthetic value of works of art, on this hypothesis, would be a function of their aptness for engaging and rewarding the attention of a person whose cognitive approach to the world is of this character.

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. The authoritarian personality. New York: Harper, 1950.

Barrett, Dorothy M., & Eaton, Elizabeth B. Preference for color or tint and some related personality data. J. Pers., 1947, 15, 222-232.

Barron, F. Some personality correlates of independence of judgment. J. Pers., 1953, 21, 287-297.

Barron, F. Threshold for the perception of human movement in inkblots. J. consult. Psychol., 1955, 19, No. 1, 33-38.

Barron, Frank. Psychology of imagination. Sci. Amer., Sept. 1958, 199, 50.

Barron, F., & Welsh, G. S. Artistic perception as a possible factor in personality style. J. Psychol., 1952, 33, 199-203.

Broverman, D. M., & Lazarus, R. S. Individual differences in task performance under conditions of cognitive interference. J. Pers., 1958, 26, 94-105.

Bulley, Margaret H. Art and everyman. Vol. 1. London: Batsford, 1951.

Child, I. L. The relation of somatotype to self-ratings on Sheldon's temperamental traits. J. Pers., 1950, 18, No. 4, 440-453.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Eysenck, H. J. The questionnaire measurement of neuroticism and extraversion. Rivista di Psicologia, 1956, 50, No. 4, 113-140.

Forster, Nora C., Vinacke, W. E., & Digman, J. M. Flexibility and rigidity in a variety of problem situations. J. abnorm. soc. Psychol., 1955, 50, 211-216.

Franck, Kate. Franck drawing completion test. Preliminary manual containing directions for scoring, norms, scoring key. Melbourne: Australian Council for Educational Research, n.d.

Franck, Kate, & Rosen, E. A projective test of masculinity-femininity. J. consult. Psychol., 1949, 13, 247-256.

Gardner, R., Holzman, P. S., Klein, G. S., Linton, Harriet, & Spence, D. P. Cognitive control. Psychol. Issues, 1959, 1, No. 4, Monogr. 4. New York: International Universities Press.

Gough, H. G. Identifying psychological femininity. Educ. psychol. Measmt, 1952, 12, 427-439.

Guilford, J. P., Kettner, N. W., & Christensen, P. R. A factor-analytic study across the domains of reasoning, creativity, and evaluation. I. Hypotheses and description of tests. Univ. Southern Calif. Rep. Psychol. Lab., July, 1954, No. 11.

Hevner, K. The affective value of pitch and tempo in music. Amer. J. Psychol., 1937, 49, 621-630.

Jung, C. G. Psychological types. New York: Harcourt, Brace, 1923.

Keehn, J. D. A factorial study of tests of color-form attitudes. J. Pers., 1955, 23, 295-307.

Knapp, R. H. n Achievement and aesthetic preference. In J. W. Atkinson (Ed.), Motives in fantasy, action, and society. Princeton: D. Van Nostrand, 1958. Pp. 367-372.

Kris, E. Psychoanalytic explorations in art. New York: International Universities Press, 1952.

MacKinnon, D. W. The nature and nurture of creative talent. Amer. Psychol., 1962, 17, 484-495.

Miller, D. R., Swanson, G. E., et al. Inner conflict and defense. New York: Holt, 1960.

Myers, Isabel B. The Myers-Briggs type indicator. Manual. Princeton: Educational Testing Service, 1962.

Osgood, C. E., et al. The measurement of meaning. Urbana: Univ. Illinois Press, 1957.

Schafer, R. Regression in the service of the ego: the relevance of a psychoanalytic concept for personality assessment. In G. Lindzey (Ed.), Assessment of human motives. New York: Grove, 1958. Pp. 119-148.

Sheldon, W. H. The varieties of temperament. New York: Harper, 1942.

Stroop, J. R. Studies in interference in serial verbal reactions. J. exp. Psychol., 1935, 18, 643-661.

Thurstone, L. L. A factorial study of perception. Psychometr. Monogr., No. 4. Chicago: University of Chicago Press, 1944.

Witkin, H. A., et al. Psychological differentiation; studies of development. New York: Wiley, 1962.

Wölfflin, H. Principles of art history. New York: Holt, 1932.

CHAPTER IV

EXPERIMENTS ON LEARNING TO DISCRIMINATE BETWEEN BETTER AND POORER WORKS OF ART

Expert judges can make rather consistent discriminations between what they consider to be better and poorer works of art. Several exploratory experiments were done on whether other people could learn such a discrimination, or could learn to improve such a discrimination.

The purpose of these experiments was twofold.

1. To study the learnability of a discrimination between better and poorer works of art, presented in pairs in the manner of the study; that is, to discover whether the discrimination can be readily learned, and under what experimental conditions. Perhaps a discrimination between poorer and better works of art is a simple consequence of exposure (under conditions ensuring careful attention) to a large number of works of art varying in esthetic quality, so that learning would result from experimentally producing this exposure. Or perhaps discrimination between poorer and better works of art is largely a product of gaining information about which works of art are considered good and which are considered poor by experts; some indication of the possibility of this mode of learning could be explored by giving experimental subjects the opportunity of gaining such information.

2. To explore the correlates of individual differences in learning -- to determine, for example, out of a number of personality characteristics measured which ones appear to predispose an individual toward rapid learning. and which predispose the individual toward slow learning or no learning at all. This second purpose can be achieved only if learning is obtained and there are consistent differences among individuals in their learning.

Procedures

These experiments were all conducted in individual sessions. The subjects were selected from those who had taken part in the study of individual differences described in the previous chapter. Certain of those subjects were invited to take part in the training experiments, and most of those invited did take part. The sessions were conducted approximately at the rate of three one-hour sessions in a week, the number of sessions for some subjects falling to two a week and for some rising to four per week. Subjects taking part were informed about the purposes of the experiment and were asked to cooperate by avoiding other influences on their judgments about art, so far as could be done without disruption of their current mode of life. In general the subjects appeared to comply with this rule, avoiding reading or deliberately entering discussions about the nature of esthetic values despite the interest that in some of them was stirred up. Usually, a subject was exposed during the hour session with the experimenter to 120 pairs of slides. Sometimes the number fell short of this, but it never exceeded this. The subject was in a laboratory room with the experimenter, seated where he could see the slides projected in front of him on the wall, with the experimenter somewhat to his rear and out of his normal field of vision.

The experiments involved four variables. Because the study was exploratory rather than systematic, and because there was interest in individual differences as well as in uniform effect of variables, these variables were not systematically studied in relation to each other. Instead there was a series of small experiments, as will be described below.

1. Feedback vs. no feedback. One group of subjects began the experiment with a series of sessions in which they were simply asked to express a judgment about the relative esthetic value of the two pictures shown them, with no information being given them about the evaluation of the pictures by anyone else. The subject was asked not to seek for information about the pictures, and the experimenter did not provide any. In contrast, another group of subjects began the experiment by being asked to predict for each pair which picture was the one probably considered the better work of art by experts, and immediately after he had given his judgment he was told whether he was right or wrong. He was, moreover, given an additional period of about five seconds in which to study the pictures again in the light of this new knowledge. The same procedure of providing this feedback was followed with the first group of subjects as a second stage in the experiment, and the course of their performance during the second stage with feedback could be compared with its course during the first stage of no feedback.

2. Initial experience with feedback vs. experience with feedback following experience with no feedback. It follows from what has been said in the previous paragraph that a comparison could be made between the subjects receiving initial training with feedback and those who received identical training after the experience of being exposed to the same pairs for the expression of esthetic judgment of their own but without feedback.

3. Easy pairs vs. difficult pairs. In the procedures described in the first two paragraphs the stimuli were easy pairs as defined in the previous chapter--that is, pairs on which no more than 2 of the 14 expert judges had disagreed with the person originally selecting the pair as to which was the better work of art. The same subjects were subsequently given training with feedback with difficult pairs, again as defined in the previous paragraph--pairs on which there was more disagreement among the expert judges but on which there was still reason to believe that there would in the long run be agreement as to which was the better work of art. There was thus no controlled experiment on the effect of training with easy pairs and difficult pairs at the same stage of training, but the general possibility of gross improvement can be gauged.

4. Low scorers vs. high scorers. All of the above procedures were intended to be used primarily with students who in the group testing of esthetic judgment had received low scores. The standard used was that out of 120 pairs in the easy test, the subject should agree with the judgment of the experts on no more than 70 pairs. Except for a very small number of subjects who agreed with the experts on fewer than 50 pairs, their scores thus fell within 10 points on either side of the score of 60 which would be expected in the long run by random response which had no relation to expert judgment. The same procedures were simultaneously applied to a smaller number of subjects who initially showed good agreement with the experts. The purpose was twofold: (a) to compare their learning with that of the low scorers; (b) to disguise the nature of the research, in order to prevent new subjects from deliberately getting low initial scores so as to be eligible for further participation.

The subjects all took part as paid subjects; but motivation for the training experiments turned out to be importantly influenced by the intrinsic interest of the activity involved and the satisfaction of learning something that many subjects felt to be in one way or another important to know. The sessions were variously conducted by the author and several assistants. At the end of each of the types of training described above, the author interviewed each subject about his reactions to it.

Findings on Learning in Low Scorers

A. Learning by Experience without Feedback

Nineteen subjects who were low scorers in the initial test of esthetic judgment began the experiment with the protracted experience of expressing an esthetic judgment about each of 960 successive pairs, with no information about whether the judgment did or did not agree with that of the experts. They were, in effect, invited to change their judgments in the course of the experience, by being told that we were interested in whether extended experience of this sort would lead to any changes in judgment.

Since it was not clear in advance whether there would be any sizable effect, we decided to guard against the unlikely possibility that a single random arrangement of the pairs would create an apparent effect. This was done by presenting the 32 pairs of magazines (each containing 30 slides) in one order to half the subjects and in reverse order to the other half of the subjects. For a subject's responses to each successive 30 pairs, the number on which he agreed with the opinion of experts was determined. This yielded for each subject a series of 32 points, to which a straight line was then fitted. A positive slope for this line would indicate that there was a tendency for the person to increase his agreement with the experts during this series of sessions. A negative slope would indicate a tendency to decrease his agreement with the experts. Of the 19 subjects, 13 showed a positive slope and 6 a negative slope. This is slightly suggestive of a general tendency toward increased agreement. When, however, the significance of the deviation of these slopes from zero is tested by t, the value of t is only 1.45. This provides no satisfactory evidence of a general tendency toward increased agreement with the experts.

A comparable analysis was made separately for the course of individual scores on the odd-numbered items and on the even-numbered items. The slopes separately determined for these two sets of items for each subject were compared in order to see whether there was consistent variation among individuals in tendency to increase or decrease. The two slopes showed a correlation of +.50, providing evidence significant at the 5% level of genuine individual differences in the slope of the line. We can conclude tentatively that there is a real difference among individuals in the changes produced in their judgments by this experience. An obvious possibility is that some individuals show an increased agreement with experts, and that others show no consistent effect. Since, however, there is no sizable tendency for positive slopes to be larger than negative (the mean of the 13 positive slopes was .071, with a maximum of .15, while the mean of the 6 negative slopes was .062 with a maximum of .13) the more appropriate inference may be that some people show an increased tendency to agree with the experts while others show an increased tendency to disagree with the experts (still others, probably, undergoing no change).

The 960 pairs presented without feedback constituted, of course, a test of esthetic judgment 8 times as long as the initial easy test, but comprised of items drawn from the same pool. Scored for total percentage of agreement with the experts, the performance of the subjects indicates that by selecting only those scoring close to the 50% level on the initial easy test we by no means eliminated important individual differences in esthetic judgment. The 19 subjects varied, in amount of agreement with the experts on the 960 pairs, from 39% up to 71%. This variation was not just the inevitable effect of steady gain or loss on total score, for the slope of the practice line was predictive of total score only to the extent of a correlation of +.23. The variation appeared to be primarily an individual difference present from the start, for the percentage of agreement was correlated +.58 with the initial difficult test and was even more highly correlated (r = +.74) with the initial easy test ddspite the fact that the subjects all had scores on the latter between 50 and 70 (except for one subject with a score of 44). For the most part, then, each subject was behaving in a stable manner throughout this experience without feedback. The tendency for some subjects to increase their agreement with experts, and for others to decrease, is a phenomenon of special interest if it is characteristic of exposure to a great variety of art, but it is not as conspicuous here as the general tendency toward maintenance of the same reactions with which one starts.

In a brief interview conducted with each subject at the end of these no-feedback sessions he was asked whether he thought his judgments had changed and in what ways. Varied extracts from these interviews may give some impression of the feelings of college students who are very intelligent but naive about art, when asked to make esthetic judgments. It is tempting to suppose that students successful in coping with the problems posed by this task would make statements radically different from those made by students unsuccessful in coping with the task. For this reason we will present separately statements by students who showed various levels of performance and change in performance as measured by agreement with experts. The reader is warned in advance, however, that there was no very regular relation between interview statements and agreement with experts.

Let us first consider the interviews of 5 subjects with high performance records. There is enough correlation among various performance measures so that these 5 include the 4 with the largest positive slope, the 4 with the largest gain on the easy test of esthetic judgment (from initial testing to a re-test at the end of these sessions) and the 4 with the largest number of agreements with the experts during these sessions as a whole.

Two of these 5 students do not claim to have changed during these sessions; one of these two feels he has not changed, the other is just completely uncertain. One of them says that in making judgments he has drawn heavily upon memories of a course in art history he took several years ago. He reports that he recognized some works or artists and inferred that the ones he recognized were the ones considered better; he also applied dimly-remembered principles of what is good in art. In short, he implies that he does not himself feel that one work is better than another, but believes that he has to some extent been applying knowledge to the task of predicting which one experts consider better; he sees no reason to think his performance in doing so has changed during the sessions, as his ability to do so derives from previous learning, not from learning here. The other subject who does not report changes makes no reference to previous knowledge of art except by implication in saying, "A few I've recognized as work by masters, but very few," and in describing some standards as having been his from the start of the sessions, e.g.:

> The attention the artist pays to detail and background... A good deal of it is just the one of the two that seems less fussy, that is more concerned with just being a work of art rather than impressing you. Also how well the color goes together.

In contrast to the first subject, this one refers to the possibility of learning something new from this experience:

> (Would you guess you agreed with the experts more at the end than at the beginning, or less, or about the same?) I would think it would be about the same, with it seeming to me I'm not changing my standards. But it's possible some things I wouldn't have been aware of at first I would become aware of. That might lead me to agree more, but I wouldn't be sure.

The other 3 of the 5 high-scoring subjects all are confident their judgments or their bases have changed, and are inclined to think the change is in the direction of greater agreement with the experts. One of these three attributes his judgments, as did one of the subjects mentioned earlier, partly to the influence of a course in art taken several years ago. He talks at some length about standards he has been using in judging and when asked where these standards come from he says,

> I thought I developed these standards in my introductory art course, and they have improved here, and these sessions have brought back what I learned at one time. (....Did the course stress comparisons between better and worse?) No, all we were concerned with was what was good in what we saw. Though I remember once he put up a poor portrait and compared it with a good one. He was primarily concerned with our acquaintance with art.

Earlier, when asked how his judgments had changed, he says,

> A specific example would be, a contrast between what you might call a modern drawing and one going back to the 17th century or something. (He is in most cases probably wrong in inferring such a difference in time.) At the beginning, I think I chose the modern more, because I thought they came closer to representing actual portraits of people. In the latter part, I began to see that actually the other portraits were better than I thought they were, in most cases. Toward the end I felt a little more confident as to judging between better and worse, because I had seen so much and had a reasonable confidence about some of my judgments, felt I had an intuition about some of them, perhaps a majority. A lot of times I noticed the names of the painters were there--Degas, Monet. Most of the time I chose the one with that name, but before doing so I actually considered which one was better and in most cases it turned out to be the more famous one. Toward the end my judgment became more what I thought was better rather than what appealed to me. I was better able to distinguish between these two.

He expresses the belief that he increasingly agreed with expert judgment, and when asked how he came to this belief he says,

> That's hard to answer. I remembered more of what I had learned, felt more confident of my decisions, and hence felt I was agreeing more with the critics.

This subject thus describes a more creative or emergent use of previous learning than did the other subject who talked of an art course; this one speaks of making use of his previous knowledge as an aid in new learning in this experiment.

Another subject similarly stresses increased confidence, but without any reference to previous learning:

> I notice my confidence is increased in making judgments. I don't know just how my judgment has been affected except in what to look for. I have formulated some opinions of things or combinations of things to look for. (Would you give me some examples?) My choices now seem to be made more on a comprehensive look at the art, whereas before on one single aspect. Now, it's on the whole context, before it might be the subject matter, or the form, or the color....If I noticed one thing striking, which didn't appeal to my esthetic taste I would rule it out and choose the other picture, whereas now I make a more comprehensive judgment. Perhaps it's that before I didn't have as much to go by, and now have more of a standard of reference.

When asked why he believes he has increased his agreement with the experts, he says,

> I feel I've just had more contact with making this type of judgment, might be better qualified.

The fifth of the high-scoring subjects is of special interest. He showed the next-to-largest gain on the easy test (changing from 70 agreements to 97 agreements out of 120) and had the fourth largest total number of agreements with the experts (628 out of the 900 pairs, or 65%). But he showed no steady gain during the 32 successive sets of 30 pairs; in fact, the slope of the line of best fit was slightly negative. These facts would indicate that the increase in his agreement with the experts came at the beginning of the sessions, and indeed might signify mostly that his performance at the initial testing was erratically low for him. Yet he is the one out of these 5 subjects who most fully and explicitly asserts that he has developed standards of judgment for himself in the course of this experience. It is consistent with the lack of improvement in his performance, however, that he asserts principally that he has developed an awareness of why he prefers one work to another, while suggesting that he may have acted on the same grounds before without awareness.

> I think I've become much more aware of what I like and what I don't like. I think I could tell you what colors I prefer more easily now than before. I've become more tolerant of different kinds of subjects, and don't let the subjects carry the full weight of whether I like the paintings. For example, religious paintings I formerly didn't like at all. Now I could enumerate, for example, the colors in it, how the forms are elaborated. I've become much more appreciative of the technique.... I find myself not liking very regular patterns, in rugs for instance; many of the designs in modern painting--I guess it's modern--dividing canvas in regular blocks, and so forth, I find don't please me........I suspect even before I didn't like paintings so well in which a pattern is very easily seen. Also vases, bowls, buildings, in which there are just parallel lines, and not very much intricacy....Before, I was just unaware of the various things one could do with a paintbrush. I've probably seen more paintings here than in my whole life before, and seeing them has made me aware.

This subject does not express great confidence in his having increased his agreement with the experts. What confidence he has is based, as far as his statement goes, on his own sense of what pleases him, rather than on external knowledge.

> I think if my standard of judgment is the same as that of those who chose, then yes, I think as my standard develops the proportion agreeing with the artists or art historians must be increasing. In early sessions much of my judgment was just a guess rather than a real judgment. (Would you guess that your standard is the same as theirs?) Yes, I guess I think it is. (Why?) I don't really know what their standard is. But if the means of expression in art is what I think it is--color, form, the mood a painter can give to his painting--I think if we're judging the same things, well, I think we are judging the same things but I'm not sure if our judgments are the same.

At the other extreme, 2 subjects out of the 19 who went through this phase of the experiment showed a decidedly negative slope--that is, showed a strong tendency to decrease their agreement with experts during the course of the sessions. During the entire series these subjects agreed with the experts somewhat more than 50% of the time; one of them had 59% agreements on the first half of the pairs and 52% on the last half, while the other dropped from 55% to 47%. They both indicate some belief that they have increased their degree of agreement with the experts, although both express it with reservation or doubt. One of them says,

> I would imagine over most of the experiment it was pretty constant--and lately I hope it increased. But I wouldn't guarantee it. (On what do you base your guess that you have increased your agreement with the experts?) It's because of the fact that I have studied it a little, and have become conscious of these other aspects.

The other of these subjects, when asked whether he would guess he has come to agree with the experts more or less, replies,

> Sometimes my feeling for a picture overrules what I thought probably the critics would say. I couldn't suppress my personal taste sometimes. (Do you suppose you did more or less as time went by?) Somewhat more, as I noticed more things to look for. (What do you think was the effect on your agreement with the critics then?) I think I probably came to agree with the critics more as time went on. (Why do you guess this to be the case?) Because with more exposure to slides, trying to pick out details which were attractive or unattractive. I've never concentrated on pictures this way before.

One of these 2 subjects cites as bases for his judgments some that might well be stated also by a person who had increased his agreement with the experts.

> At the beginning I had very few bases other than prejudice--what appeals to me. Photographic quality was one thing that appealed to me. Though I still like representation of reality, I look for other elements too--balance, contrast of color, and such elements that some people say are what makes good painting. My conscious criteria are more explicit, though they may be the same criteria.

It is noteworthy, however, that he cites such considerations only as what he thinks some people may say make good art, not as elements in works of art that make them really seem better to him or excite his interest more. The other of these 2 subjects formulated criteria that a critic would presumably regard as irrelevant to the quality of the art.

> When I recognized something as the work of an old master, I chose it even though in many cases I felt it was not really the better. And many of the choices between women--I chose the slimmer one because of a distaste for the other, except where I recognized the other one as Botticelli or something--he likes fat women..... I preferred a blue or green background, rather than red or brown or gold background. The 2 pictures were usually opposed in this way, so consistently it might be intentional.....In sculpture I usually chose what I thought was the old work of art.....Also, the ones that

> looked more realistic, the ones that looked more like a photograph than like a painting--I would choose these....In portraits, I usually chose one with more light on the face rather than one with half of the face in shadow.

Another group of subjects that may be interesting to consider consists of those who showed a consistent tendency to select as better the picture the experts consider poorer. The three people at the extreme in this respect, instead of the 50% chance value of 480 agreements with the experts out of 960, had 377, 388, and 390 agreements. The middle one of these showed a fairly high tendency to increase agreement with the experts during the series, while the other two had nearly zero slopes. None of them changed appreciably in the test, from the initial test to the repetition of the easy test at the end of these sessions. The middle subject gained 5 points from 44 agreements to 49, the first one went down 4 points from 55 to 51, and the last one increased from 52 to 55. All these scores are below the 50% chance value of 60 out of 120.

Of these three students, one believes that he has increased his agreement with the experts during these sessions.

> I suppose the more I've looked the more I've been able to pick out which were the best....I think I've come to agree more with the experts as time went on.

When asked to formulate criteria he used or came to use, he replied as follows:

> I don't think I had any to start with--just the ones I liked. But after a while I began to look for color combination, the expressions on the faces where there were people. (What did you judge to be better and what worse in these cases?) It's hard to say. I tried to look into the faces and see if they were in accord with the theme--if the mother was interested in the baby, for example, and vice versa. (And how about the color combinations?) Some of them just looked bad, that's all. I tended to pick those that seemed more natural.

The other 2 subjects had no such belief in their own improvement. One said he saw no reason to believe his judgments had changed, and when asked about agreement with the experts said,

> It's hard to say, because I've never known up to now anything about what the experts agree on.

When asked about the criteria he was using or had come to use, he replied:

> No, I think I've kept my ideas roughly the same. I usually look for composition, colors and pose and I don't think my ideas have changed. Some of the choices I've made are the choices of the lesser of two evils. Sometimes they're both quite good. (What were your bases of choice as far as composition is concerned?) The framing, let us say, around the outside--for example, an arch which will direct attention toward the center. Or in landscapes, a road or line of trees which directs your sight. Or a town in the distance, with trees in the foreground so you get visions of depth. Your framing has a lot to do with it. In color it's hard to say, You have to have some contrast. Two vivid colors together are not so good. Some of Rembrandt's dark colors are good because they leave a lot to your imagination. As for pose, some activity is good, and a person turned slightly aside is better than looking straight forward. As far as the modern art things go, I'm afraid you'll have to leave that blank--except that the ones that convey a feeling of happiness or have lots of colors are better than the ones that are black and have very little color. Oh yes, another thing about composition--where there is a bunch of people you can have it too crowded, not centered on the important individuals in the scene.

A sense of confusion about subjects' self-report in relation to their performance is an appropriate impression to gain from these interview abstracts. The interview protocols were rated for four variables: sense of change in performance, esthetic relevance of criteria stated for choice, extent of dependence upon these criteria, and self-estimate of increase in agreement with the experts. There was a consistent tendency for these variables to be positively correlated with the total number of agreements with expert opinion (correlations, in order, of +.26, +.38, +.23, +.28, with N of 17 or 18), and nearly so with gain on the easy test (correlations of +.26, +.42, -.07, and +.22). But the relation to slope of performance was very irregular, with correlations of -.25, +.30, -.03, and -.37. These correlations are not large for the kind of relation involved, and the one closest to relating two expressions of the same thing--real increase in agreement with experts, and the subject's belief about whether he has increased--actually has the negative value of -.37. For the other aspects of training to be presented later it is also true that the relation between performance measures and interview variables is low. For demonstration of this point we will rely upon this first instance and in the later presentations will give interview data principally for what they may reveal in themselves rather than in relation to performance measures.

B. Effect of Easy Training with Feedback

For 21 subjects the first experimental procedure was to ask them to judge which picture in each pair is considered better by experts, and to tell them immediately after each judgment whether it was right or wrong. In response to this procedure, every one of the 21 subjects acquired or improved a discrimination between the better and poorer works of art. In the way described in the previous section a straight line was fitted to the succession of points representing the number of correct judgments in each successive magazine of 30 pairs. For every one of the 21 subjects, the slope of this line was positive, and indeed for no one was it close to zero. The mean of these slopes was very significantly different from zero ($t = 15.0$).

The results were similarly decisive, though not quite so extreme, for the subjects who had this experience with feedback following the experience without feedback that has already been described. Of the 19 subjects in this group all but one showed a positive slope--that is, increase in proportion of correct judgments through the series of sessions. The one subject who was an exception had a slope just barely negative, and one additional subject had a slope that was just barely positive. The group as a whole showed decisive evidence of learning a discrimination. The t for the difference between the mean slope and a slope of zero was 7.85.

If there is such clear evidence of learning a discrimination between the better and poorer works of art, we may then ask how good a discrimination is established? Here are intelligent students, generally very responsible to their commitment as paid subjects and many of them highly motivated to profit by the learning opportunity offered them, who make about 1000 discriminative judgments each followed by immediate knowledge of results. Their experience of this sort in the experiment, we may infer, is considerably more pointed and more directed at establishing a clear discrimination, than that which the experienced student with a prior interest in art will normally have had in the past. The person with an interest in art will, of course, have seen in galleries and books pictures he may assume are in general considered to be of high esthetic quality, and he will have seen in magazines, newspapers and gift shops many works of art that he may infer are not considered to have high esthetic quality. But rarely will he have seen works within a general category presented side by side and have been informed about which one the experts consider better. Particularly is he unlikely to have had such comparisons made for him between works which resemble each other in subject matter and style, so that the elements involved in the esthetic judgment can become especially conspicuous.

The most satisfactory way of judging the quality of the discrimination achieved by the subjects is provided by the repetition at the end of the training of the easy test and the difficult test given before training began. The subjects were now asked to apply what they had learned in the training sessions, and to try to guess for each pair in the test which work was the one that experts consider to be the better. The amount of change in test score will first be considered, and then the absolute level reached.

Of the 21 subjects for whom this was the first experimental procedure undertaken, every one showed a substantial gain in score on the 120-item easy test of esthetic judgment. The mean gain was 34 items; the smallest gain was 21 items, and the largest was 45. The amount of this gain that should be attributed to a statistical artifact (regression, resulting from the group's being selected for low initial score and then being retested) may be estimated as probably no more than about 7 items, which was the mean gain of the other group of subjects when they were retested after the exposure to 960 pairs without feedback. The evidence of gain was equally unanimous on the difficult esthetic judgment test, though the gains were smaller. The mean gain was 13 items, the maximum 19 and the minimum 3. Since scores on the difficult test had no influence on selection of subjects, this gain may all be regarded as genuine.

The subjects who began the experiment by making esthetic judgments without feedback were, as has already been indicated, given the easy test again after that experience. Their performance at that point may be compared with their subsequent performance on it after the training with immediate feedback to estimate the effects of this training. There was a large general improvement which averaged 23 points but was not shared by everyone. Three subjects had already gained 20 points or more on the easy test after the no-feedback experience (as compared with the initial test). Two of them now showed only slight additional gains (2 points and 6 points), and the third actually dropped 8 points. The subjects who began by judging without feedback had the difficult test only in the initial group session and at the end of the training with feedback. Their mean gain on it was 16, slightly more than for the other group. But again there was not perfect uniformity. One subject received the same score as he had initially, and one subject fell by one point.

How accurately, then, can these subjects predict the judgments of the experts, after being trained to do so? The mean score of both groups of experimental subjects on the post-training test is 92 items right out of 120; the highest scorer reaches to 104, and the lowest only to 78. If we compare this record with that of all 138 subjects on the initial testing, we find that the mean score after this training corresponds to the 92nd percentile of the distribution, and the lowest score after training corresponds to the 61st percentile of the distribution. The trainees have thus acquired a skill which is decidedly above the average of that developed by college students in the normal course of events. On the other hand, the training directed at development of just this skill does not lead to developing it beyond the degree to which it is already found in some students. Out of 138 subjects who took the initial test, 4 had higher scores (108, 106, and two 105's) than any of the low scorers attained as a result of training, and these 4 obtained their high scores under conditions in which they were asked to express their own judgments rather than being asked to predict the opinions of experts against which their judgments were scored. These high scorers had all had a good deal of experience with art, to be sure, but not experience of this particular kind.

On the difficult test of esthetic judgment, the average effect of the training is almost as great, but there is more variability. The mean score after training (73 agreements out of 120) falls at the 87th percentile of the total sample of 79 students who took the difficult test in the group sessions. The lowest post-training score was 60, which falls at the 15th percentile of initial scores; altogether, 9 of the 39 post-training scores fall below the 61st percentile, whereas on the easy test none did. At the other end of the distribution, the top score after training is tied (at 85) with the next-to-highest initial score, and the single initial score that was higher is only 87.

It is clear, then, that such training as this leads almost all the low scorers to discriminate with an accuracy that rivals the initial discrimination of high-scoring students. It does not, however, enable any of the low-scoring students to end up with discrimination more accurate than the high scorers had at the beginning.

The interview conducted at the end of this training was much more thorough than the brief interview conducted after the sessions without feedback. One portion of the interview reveals strikingly the variations among people in the meaning that this training experience had for them. The experimenter posed the following question:

> (Now I would like to describe to you a couple of hypothetical situations, which are not true, to ask how you would react to them. First, suppose you were talking about this experiment with someone, and the other person tried to make you believe that the information you were given here about the judgments of experts was all

> false, in the sense that it had just been made up at random--as though we had flipped a coin for each pair to decide which picture we would say was the one the experts consider better. What would you say to that? ....Now the other situation I want to pose to you is this: Suppose this other person were to try to persuade you, not that the information about the experts was made up at random, but instead that it was all exactly the opposite of the truth--that is, that the pictures we told you the experts considered better were really the ones they considered poorer and vice versa. What would you say to that?)

At one extreme, there were subjects for whom a consistency in the information given them had been so apparent that there could be no possible doubt on the first question asked. For example:

> That's an interesting question. I thought of this myself. But after a while you see a pattern begin to emerge. And even a person like me with no art training becomes aware there is a pattern. I may not know what the criteria are, but I am very well aware that there are criteria.

Much the same point could be made in response terms rather than perceptual terms, by the student's saying that he had used consistent methods of choosing, and since they worked, the choices could not possibly have been random:

> I would deny it, because I know better. Although in some cases I disagreed with the judgments, they were few. I would use the standards I used in making my judgments to deny that it was a random choice.

Experience outside the experiment could also be called on, for evidence of consistency:

> I couldn't agree with it, because there has been a definite system about what is good repeated again and again. I could tell for sure that coins hadn't been flipped. Also I'm taking History of Art, and it certainly seems to me that the ones I had seen there were the right ones here.

The reference to outside experience was a frequent help in response to the second part of this question, about the possibility that the information had been completely reversed. Many students pointed out that they could recognize some pictures as acknowledged masterpieces, and that these were practically always the better of the pair according to the information given them. They were then asked what they would say if it were argued that the experiment had been carefully rigged so that pictures that would be recognized as acknowledged masterpieces were correctly said to be the better ones, but that in all other cases the opinion of experts had been represented as the opposite of what it really is. A few subjects indicated they would then have no argument to offer in reply. Some, however, extended their feeling of consistency from pictures to which outside information applied directly to many other pictures in the series.

> I couldn't be persuaded of it. There were definite similarities between the ones I recognized and knew as good from History of Art and the others reported as good, similarities which I could see.

Some students used outside information, or information gained from artists' signatures seen in the course of the experiment itself, to confirm an observation of consistency already made independently:

> I spotted Renoir paintings as good ones without knowing them yet, just from the theory I had developed from looking at the other pairs. There were other examples of this, too.

Another argument made by some students was that the consistency was much greater in what was reported to be good than in what was reported to be bad.

> There was no consistency in what was not the choice of the critics, as far as I can see. The dissimilarities were more great than the similarities. So if there was no consistency there, there could be no standards. (This is an explanation of why he believes the information given would not have been the reverse of the truth.)

At the other extreme there were students who report no experience of consistency at all.

> I would be inclined to agree with that very strongly, except for the few I knew to be by the old masters, and those were always correct. But I actually thought that this might be the case. (i.e., that the supposed information had been made up at random)..........I could be led to believe this without much trouble, since I couldn't learn to perceive what it is the critics like in these pictures....... It wouldn't be hard to believe they were telling me the exact opposite of the truth--that the purpose of the experiment might be to see whether you could change a person's innate esthetic judgment....... I have no conclusive evidence to draw on, since I haven't been able to seek outside sources. The works of art I recognized, however, were always the better.

It was even possible to reconcile recognition of some pictures with a belief that there was no consistency.

> (Misinformation at random?) In most cases my ignorance of art is enough so I'd believe it. Occasionally a piece of sculpture or painting would come up that I had seen. Except for those, I don't think I'd ever have known the difference. (Information reversed?) Again, except for the few pictures I recognized I'd have to believe it. Of course there are experts who disbelieve in the masterpieces and it could even be rationalized on that. My ignorance of art is such that I'd believe just about anything I was told in here.

In another part of this interview each subject was shown 10 pairs different from, but of the same kind as, those he had been seeing during the training sessions and was asked to predict for each pair which was the one the experts consider better, and then to explain the reasons for his choice. No interview of this sort had been conducted at the beginning of the sessions, but it seems reasonably certain from the combination of changes in performance and the nature of what some subjects said during this interview that their reasons represented the real extraction of some generalizations from their experience in the training sessions. For illustration we will use the last of the 10 pairs that were shown to each subject on this occasion. The picture on the right was a Picasso painting similar to those of his blue period but predominantly in pinkish brown. It shows a lean woman bending over her ironing board. On the left was a rather Cubist picture by Rattner, which might superficially be taken to resemble closely another of Picasso's styles. It shows a woman probably cutting bread. A great effect of the training can be illustrated by the response of one student who at the initial interview in the individual session before any training was started had described himself as having no background in art beyond the eighth grade, as not reading art books or sections on art in magazines, and as having been to art galleries only on occasion with a friend or relative. He had then described art in general as "a lot of stuff that doesn't interest me in the slightest--unnecessary." In this pair he chose the painting on the right and said:

> Again, as in a lot of the pairs, there is a lot about both of them I like. The one on the right is more pleasing in total because of the similarity of color. You can almost feel what it is the woman is doing--ironing--you can almost feel her poverty, and that she is working sixteen hours out of the day. The stoop on her shoulders and back, the leanness of the body. In the picture on the left you have to look pretty hard to see just what it is. But my positive attraction to the one on the right is the principal reason. The one on the right just seems real, more easily felt.

Giving this sort of reason could even be combined with picking the picture experts consider to be poorer, and maintaining independence of judgment in doing so despite at least a tentative recognition that the one on the right is by a well-known artist who is much respected, as in the following case from a student who chooses the painting on the left.

> I'd pick the one on the left. But I would also say I like the one on the right too, very much. I think the one on the right is done by Picasso. The whole painting in many ways possesses just one color, just variations on the basic color. A lean figure standing over the table. Very motionless for a painting. I like it so much I find it hard to say it's inferior to the one on the left. I like the one on the left primarily because of the use of color there. It too has emotional quality; the greens of the woman's dress, her face, present a strain as she cuts her cheese. Even though it has contrasted colors, they don't seem out of place; they seem balanced. One reason is that it's hard to mark a line exactly where one color ends and the other begins; perhaps that's necessary if you have a lot of colors.

When explaining their choices in this part of the interview, of course, some subjects show the effect of their experience in the training sessions without at all giving kinds of responses that art experts might be expected to give. These are explanations in which the subject views the task as entirely external to him, reports having obtained from the experiment clues about what it is the critics say is good in art but with no indication of personal meaning for him. For example:

> It's the one on the left. It seems to be along the style of Picasso, though I don't believe it's his. The colors are too garish to suit me.

Or for another example:

> The left. I feel I should pick the one on the left, though I don't think I personally like it that well. The reason I don't pick the one on the right is--the right looks very well painted, but the subject seems, that face just wouldn't be accepted by critics. Then again perhaps it's too modern looking; the one on the left gets away with it because it's on an abstracter level.

In the remainder of the interview the subject was asked what he felt he had learned from the training sessions, what the general bases for his judgment of different kinds of pairs was, and how his attitude toward or interest in art had been influenced if at all. In general, the experience seems to have had a considerable impact on most of the subjects. This is perhaps least so for those subjects for whom the reaction to the pictures remained largely external--that is, for whom the task remained one of identifying clues from which they could make a prediction correctly about somebody else's opinion, without having come to share the opinion in any way. One student, for example, describes his bases for judgment almost entirely in terms of clues which the experts would probably consider to be irrelevant. He emphasizes two particular clues. One is that if one work looks older and the other looks newer, he finds that the one that looks older is likely to be considered better by the experts. This is especially true, he says, for sculpture, where if one is chipped or damaged it is almost sure to be the one judged to be better. The second clue he emphasizes is that if one work is less representational than the other it is likely to be judged better; this is particularly helpful, he says, in judging painting. The only indication he gives of a positive interest in the experiment, other than the fact that he has tried very hard to solve the problems posed by it is his saying, when asked about effects on his interest, that the experiment probably has increased his interest in art.

> It's a matter of having your own opinion reinforced or not reinforced. Mine were quite unreinforced, and this has stimulated my desire to know why this is so, why my opinions are wrong in the opinion of the experts.

Another student gives a greater variety of specific clues that helped him to be successful in predicting the opinion of experts. In statues the more ferocious ones were preferred. In Indian masks the thinner ones were preferred. In abstract paintings, ones with black were considered better. In portraits, the ones that looked least like a photograph. In still life paintings, it was an advantage for a painting to be less lifelike and to have pale colors and a muddy complexion. He also mentions a personal clue cited by others as well:

> I found that if I chose the one I liked least I would often be right. My esthetic appreciation before was based mostly on *Saturday Evening Post* covers. If a picture looked like a *Saturday Evening Post* cover I voted against it. Or, in general if I liked it I went against it....But it's been a very rewarding experience, it has made me much more aware of art. I even penetrated into the Yale Art Gallery for the first time in the three years I have been here.

For a number of students, what they learned had much more personal meaning. At the extreme some students said their choice of a major in college, or conceivably even their choice of a career might be altered. One example of personal meaning has been selected for presentation at length--not because it is extreme in intensity of personal meaning, but because it is a case where the evidence seems good that the meaning of art for the speaker derives almost entirely from this experience. In his initial interview he reported that he has had no art training he can remember since kindergarten, has had no artistic hobbies except for drawing some airplanes and cars around the age of ten and eleven, sees the art section of Time magazine while leafing through it but does not read the text of this section, and has never been in an art gallery except for several visits to the Yale Art Gallery with a girl friend who wanted to go there. When asked at the beginning of the post-training interview what he has learned from these sessions, his reply is the following, in which series of dots represent pauses, not omissions:

> I've learned something about form. Rather than have it cluttered or ornate, it should be blended into a whole, should be a unit rather than a collection of parts....As for the paintings: a skillful, exact execution is not always so good because it leaves nothing for you to see for yourself. It says everything rather than suggesting. .....Quite frankly, I could not see by what criteria they judged statues. While I could pick out the right one, I could not tell how they did it. It seemed to me that the better one always lost....Color also can be used as a part of a form. You should not have great clashing blobs, just as clashing shapes; they should blend and complement so no one dominates. They should work with each other to produce a coherent whole....One side effect is that I have reached a point where I could go to an art show and pick out something I like for my room--which I fully intend to do....(Anything else?) Possibly something about judging art works as was the object. By comparison at least, I have been able to find some standards by which I can judge--know more or less the kind of thing I like.... (What relation does it have to what the experts say?) We pretty much agree on architecture and most paintings--except perhaps sometimes old paintings. We agree on abstract sculpture--the sculpture I was referring to before was statues. The Chinese art I like very much, and I was completely in agreement on that--Oriental art in general. Photography we got together on pretty well too. I guess there was some disagreement with them on abstract paintings. I couldn't really tell awfully well just how they were able to reach their conclusions. When I got an abstract painting right it was more from a feeling that I just thought it was good, than from a conscious evaluation by any standards--I just thought it was good, couldn't tell why.

C. Effect of Training with Difficult Pairs

The first group of subjects, those who had previously had training only with feedback with the easy pairs, took part in training with the difficult pairs at a time when only about 500 of these had been assembled. The other group of subjects who had had the experience without feedback prior to the training with feedback with the easy pairs were subsequently given training on difficult pairs when over 1000 of these pairs had been prepared. The procedure was the same for both groups. They were told fully the nature of the difficult pairs--that is, that on these pairs it was probable in the long run that experts would agree as to which work was the better, but that there had been sufficient difference of opinion to make this doubtful in some cases, and that in any event the discrimination was more difficult even for experts. They were then asked to estimate for each successive pair which was the one the experts considered better and were then immediately told whether they were right or wrong.

In both groups the number of correct judgments tended to decline. In the first group, 12 out of the 19 subjects who went on to the difficult training exhibited such a tendency. When the magnitude of this negative slope was taken into account and the difference of the mean slope from zero tested by t, the value of t was only 1.52, indicating that a consistent tendency toward a downward slope could not safely be inferred. In the second group 10 out of the 15 subjects showed a downward slope; here the t was 2.12, giving evidence significant at the 5% level of a general tendency downward. Since the two groups were subjected to somewhat comparable experience directed at testing the same question even though the amount of experience was less in the first group, it seems reasonable to combine the evidence by pooling the probability value. When this is done (by the method described by Winer, 1962, p. 44), the overall evidence of a consistent tendency toward a downward slope is significant at the 5% level (Chi-square = 9.61, d. f. = 4).

The effect of the difficult training on test performance, on the other hand, showed no significant trend. On the difficult test, 14 subjects raised their score on re-test after the difficult training and 17 lowered it; 2 had the same score as on the previous administration. On the easy test 15 improved, 17 declined, and one had the same score as before.

The impression subjects stated in an interview at the end of the difficult training, about whether their agreement with experts had increased or not during the sessions, varied widely. The variation was almost entirely unrelated to the course of their actual performance despite the constant feedback; the correlation coefficient between the two was only +.14. It should be recalled that the measure of change in actual agreement had no internal consistency, and lack of correlation with another variable is hardly surprising. The course of performance in different subjects did not vary greatly (though the absolute level did) and was close to a rather baffling lack of change. Faced with this experience, chaotic in comparison with the obvious improvement felt by almost everyone in the preceding part of the experiment, different subjects reached very different conclusions about it. At one extreme there was at least one student who was confident he had improved during the difficult training itself, and also in the difficult test after the training.

> Oh yes, surely. But the increase was much slower than in the previous series. The criteria involved were much more subtle and complex, so you can't go ahead by leaps and bounds. (How about the test of difficult pairs?) Yes, I thought I did better on it, definitely, and I just flashed through some of them. They were like old friends.

At the other extreme, some subjects thought they had not improved by either measure, and might even have gotten worse.

> I don't know for sure. When I started off the series I seemed to be doing better than before. Later on, I don't know. I got some ideas but from the results I may even have gone down a bit...Perhaps there was progress in that I came to a decision faster. (How about the test of difficult pairs?) I don't know; I think I probably did about the same as before.

In this interview subjects were also asked what they felt they had learned during the difficult sessions and whether they thought the standards they had formulated for their judgments during the easy training still held true with the difficult pairs. There was a great deal of variation in response to these questions too. Some subjects felt that they learned a good deal from the difficult series, and that this learning was a continuation or extension of the learning during the easy training. For example,

> I think I've probably gotten more out of this last series--a better sense of balance and proportion, color, composition of colors. I just get a feeling that it is this one, and I find myself agreeing with the artists more than at the beginning....I actually didn't see much difference between the harder and the easier. Actually I felt surer of myself, more confident in my judgment. A lot of these seemed much easier to judge than in the first series, but probably this is due to training. I have some idea of what the judges were looking for, in color composition etc...(So generally the generalizations you formed before still seemed to work?) Yes, I started to form them in the previous series and have substantiated them and made them firmer. I've gotten to see things faster.

One subject feels that in the difficult training he has made automatic a global evaluation which does seem to be related to the separate criteria he had formulated before.

> I don't think I've really learned anything in particular. I've learned that I have developed a sort of instinct for picking pictures....I look at two pictures and I get a feeling, I just know, that one is better, and I can't pinpoint why. And when I try to think why, I may lead myself astray. It's the sort of thing that if I try to concentrate on it, it eludes me, while if I let it, it

> seems to sort of come by itself. (Did the criteria you were using before seem to hold true still?) They seem to have held up except that I couldn't make a definite decision with them because there was such a fine edge between the two. Whereas the instinct would just pick one out, and usually it was right.

At the other extreme an occasional student actually felt that his performance became worse during the difficult series.

> I think I became more confused. All those simple artificial technical details we picked up were lost, because these were closer and there were so many exceptions.

The report that simple clues or keys developed in the easy training no longer worked was sometimes coupled with a feeling that a skill had been retained that can lead to correct judgment.

> It sure was a lot harder. The main thing I learned was that a lot of things I learned before weren't true. I got to using fewer and fewer rules. You somehow can tell. There isn't one thing you can pick out and say, if a picture has that it's right.

It is tempting to suppose that a belief in the continued validity of the criteria formulated earlier would be found in those who had formulated criteria which seemed esthetically relevant, and that a frustrated conviction that the criteria no longer worked would be found in those who had formulated esthetically irrelevant criteria. A test of this can be made from ratings made on the interview protocols. One rating of the interview after difficult training was of just this variable: the extent to which the subject felt his previously formulated criteria had held up. Two relevant ratings were made of the interview after the easy training; the extent to which the reasons given for choices in the 10 pairs included in the interview were judged to be esthetically relevant, [illegible] view stated esthetically relevant criteria. These two ratings, correlated +.53 with each other, were correlated only +.17 and +.09 (N = 32) with the relevant rating of the interview after difficult training. A single rating of the extent to which the interview after non-feedback experience included esthetically relevant criteria was correlated +.42 with the difficult-training rating; but the number of cases is so small (N = 13) that we cannot be confident of any but a chance relationship even here. Belief that the standards previously formulated still held true was certainly, at any rate, not closely dependent upon the esthetic relevance of those standards.

Findings on Learning in High Scorers

Only 9 subjects who on the initial test showed a high degree of agreement with expert judgment were used as subjects in individual sessions. Their initial scores on the easy test varied from 88 up to 106 agreements out of 120 pairs.

Only 4 of these initially high-scoring subjects were exposed to the protracted experience of making large numbers of esthetic judgments without feedback. The slopes of their practice lines do not deviate much from zero; two are slightly upward and two are slightly downward. There is no evidence from the course of their performance in these sessions themselves, therefore, that they are any more consistently prone than the low scorers to increase agreement with expert opinion as a result of such exposure to pairs of pictures. Findings obtained with the easy test do, however, suggest such an increase. These subjects had originally been recruited for the training experiment because of their high scores on the initial easy test. When this test was administered to them again after the non-feedback experience, therefore, the statistical artifact of regression would make for their receiving lower scores. Contrary to this expectation, every one of the 4 showed a higher score than before. One rose appreciably from a score of 90 to 97; the others rose minutely, from 105 to 106, from 88 to 90, and from 91 to 92. This regular evidence of an increase may, however, not be related to the experience with 960 pairs in the non-feedback sessions. Perhaps high scorers would be able to improve their scores on re-test without any intervening experience; no experiment has been done to exclude this possibility.

Viewed as a protracted test of esthetic judgment, the non-feedback sessions do suggest that on the initial test the high scorers had somewhat higher scores than they might have had on another day or on another comparable set of pairs. Every one of the 4 agrees with the experts on a somewhat smaller percentage of pairs during the non-feedback sessions than he had in the initial test. The one with the highest score on the initial test [illegible] agreement down to 83%. The next highest drops only a point, from 76% to 75%. The third also drops only a little, from 75% to 72%. The lowest of the high scorers drops much more substantially, from 73% to 62%. The 4 subjects fall in the same rank order of scores for the initial test and for the non-feedback sessions.

The same 9 high-scoring subjects went through the easy training with feedback, the same procedure described under that label for the low-scoring subjects. Five of these had not previously had the experience without feedback and 4 of them had. The slope of the best-fitting straight line for the score on successive magazines of 30 pairs was positive for 8 out of the 9 subjects, and for the remaining 1 it was just barely negative. The mean slope differed from zero by only about half the magnitude found for the low-scoring subjects--not a surprising finding, of course, since the high-scorers start from so high a level--but the tendency toward a positive slope is significant at the 2% level ($t = 3.22$).

Improvement on the easy test as a result of the training with feedback was uniformly shown by all of the high scorers. Gains varied from 3 points to 9. There was no difference between the 4 subjects who had previously had sessions with feedback and the 5 who had not, even though the **former** group have their final test score compared with a pre-training score which has already increased above their initial score in the group session. The mean gain for those who had had the previous non-feedback sessions was 6.25 points and the mean gain for the other group was 6.4 points. The over-all gain is highly significant even by a simple sign test.

Improvement on the difficult test occurred less uniformly. Two subjects decreased their score on the difficult test, one by 9 points and another by 4. The other 7 increased their score by amounts varying from 3 to 13 points. For all 9 subjects together the mean change was 4.4 points, a deviation from zero which is not statistically significant (t = 1.7).

The level of performance on the easy test attained by these subjects after training is very high. Five of the 9 have higher scores than the top-scoring student out of the 138 who took the initial test, and another ties his score of 108 agreements. The remaining 3 with scores of 99, 101, and 102 bring the mean down to 107, just below the highest initial score of 108 out of 120. The lowest post-training score stands at the 93rd percentile of the initial scores. On the difficult test, of course, the absolute level of scoring is much lower, but the relation to the initial distribution of all students is similar to what it is for the easy test. The lowest post-training score of these high-scorers is at the 92nd percentile of the original distribution, and the mean is at the 97th percentile. Only 2 out of the 9 subjects, however, achieve a higher score than the highest initial score of 87.

The high ability to predict the judgment of experts shown by some of these subjects on the post-training test is not dependent upon their having seen the test items earlier in the study. They reach a similarly high level of performance on the training pairs in the last 5 sessions of the training one subject averages 112 judgments correct out of 120; another averages 111, one 109, two 108, and one 107. The others average 104, 103, and 94 correct.

Eight of the 9 high-scoring subjects were trained with the difficult pairs with feedback. All 8 had the longer training that was shared by 15 of the low-scoring subjects. Of these 8, 3 showed a negative slope--that is, a tendency toward decreasing proportion of agreements with the experts--during the course of the difficult training. The other 5 showed a positive slope. The mean slope of all 8 was positive but not significantly so (t = 1.02). The suggestion is of a positive slope on the average, but the most one can safely conclude is that there is no evidence of a general tendency toward a negative slope as in the low-scoring subjects. The experience of being exposed to more difficult discriminations, then, apparently does not confuse and disturb the established standards of the people who are potential high scorers before the experiment begins.

Comparison of test scores before and after the difficult training leads to the same conclusion, that in these subjects the difficult training produces little or no consistent change. On the easy test 2 subjects improved (by 1 and 3 points), 3 declined (by 1, 2, and 5 points), and 3 stayed the same. On the difficult test 4 improved and 4 declined by amounts that exactly balance each other.

The absolute level of performance reached on the difficult test suggests that even for these students who are originally high scorers it is difficult to gain much in sharpness of discrimination by application, to the difficult pairs, of the training methods used in this experiment. At the end of the difficult training not a single subject surpasses the initial top score of 87 on the difficult test; the 2 who had surpassed it after the easy training have dropped back in score, and no one else has risen that high.

Individual Differences in Response to Discrimination Training

For each kind of training experience, in each of the two kinds of subjects (high scorers and low scorers), and for groups of low scorers separated according to kinds of experimental treatment, the slopes of the straight lines of best fit were calculated for the total performance on successive magazines to determine the results mentioned above. In addition, however, results were treated separately for odd-numbered items and even-numbered items in order to test whether individual differences in the slope were consistent. The slope measure obtained for odd-numbered items was correlated with the slope measure obtained for the same persons on even-numbered items. Only 2 of the odd-even correlations so obtained gave significant evidence of any consistency at all. Both of these measures pertain to the low-scorers who were exposed first to the experience without feedback. Individual differences in slope during the initial training without feedback showed a split-half correlation of +.50,[1] significant at the 5% level as has been mentioned earlier. For the same subjects the split-half correlation of slopes during the subsequent easy training with feedback showed a value of +.66, significant at the 1% level. These two significantly consistent measures—trend of individual performance during sessions without feedback and during sessions with feedback--appear to have a slight negative relation to each other. The correlation between the two is, however, only -.21, a statistically insignificant value which may be regarded as giving no satisfactory evidence of any relation at all between the two measures.

---

[1] In the previous chapter, the split-half correlation coefficients were corrected by the Spearman-Brown formula to give the best estimate of the reliability of the whole, the reliability generally being high enough so that the significance of its difference from zero does not come into question. In the present section, the question is whether there is dependable evidence of any reliability at all; for this purpose the uncorrected coefficient is required.

Most of the other split-half correlations were close to zero on either the positive or the negative side, and they gave no suggestion of consistent individual differences in the effects of the training procedures on the discriminative judgment. The largest of these other split-half correlations was actually negative: -.37 for the difficult training with the first group of 19 low scorers. Others were -.15 for the difficult training with the second group of 15 low scorers, and +.17 for easy training with feedback in the first group of 21 low scorers. The coefficients for the training of the separate groups of 4 or 5 high scorers have not been computed because of the exceedingly small number of cases; coefficients for the pooled groups are very low.

Where the individual subjects did not reliably differ from one another, it is futile to inquire into the correlates of individual differences. Among the slope measures, therefore, only the two which have reliability as measures of individual differences can be made the subject of such inquiry. This inquiry turns out to be disappointing. The number of statistically significant correlates of these two slope measures, when related to the personality variables available, is smaller than might be expected by chance; and the few that are found do not form any coherent pattern. Thus the attempt to use the training experiments as a source of information about individual differences in esthetic trainability was not successful.

SUMMARY AND DISCUSSION

The study reported in this chapter investigates the trainability of discrimination between works of art which experts consider to differ in esthetic value. The stimulus materials consisted of the paired slides whose selection was described early in Chapter III; each pair consisted of two somewhat similar works of art, of which one was considered better esthetically by the person selecting the pair. The pairs fall into two groups: Easy Pairs on whose esthetic difference 14 expert judges had agreed with no more than 2 dissents; Difficult Pairs on whose esthetic difference more than 2 of 14 expert judges had disagreed, but for which the pattern of expert judgment still suggested the probability of long-run agreement with the original selector. From each of these two groups some pairs were used for testing (the same pairs used in the tests of esthetic judgments reported on in Chapter III), and other pairs were used in training. The training was conducted in individual sessions as were the tests which followed training.

One purpose of these training experiments was to study the personality correlates of esthetic trainability. With respect to this purpose the experiments were unsuccessful. In general, the individual differences in trainability were not reliable; where they were reliable, they showed no meaningful pattern of relations with other characteristics of the subjects. The numbers of cases involved in these training experiments are too small to justify any very definite conclusion from this lack of reliable results on individual differences.

The second purpose was to study, in general rather than with respect to individual differences, the learning of a discrimination between "better" and "poorer" works of art. Here a number of results were obtained, which will be summarized in relation to the four variables described in the introduction to this chapter.

1. Feedback vs. no feedback.

Some of the initially low-scoring subjects began the experiment by making judgments and receiving no feedback. Others received immediate feedback consisting of information about whether each judgment agreed with that of experts. The first subjects also went through this procedure after the no-feedback sessions had been finished. Training with feedback led almost uniformly to progressive improvement of performance in judging which work in a pair is considered better by experts; only one subject out of 40 failed to show a positive slope when a straight line was fitted to his successive scores. When measured by test performance before and after training (a less satisfactory measure since the subjects were picked for low initial test score), all but one subject again showed improvement in performance. The level of performance reached on the post-test by these initially low-scoring subjects was high in the distribution of initial scores, its mean falling at about the 90th percentile of that distribution, but not one of them exceeded the spontaneous agreement with experts found at the outset in the initially highest-scoring students.

The experience of judging the same pairs without feedback, on the other hand, led to no such large and uniform increase in agreement with experts. This contrast is not surprising but is well to have clearly established: an intelligent and well-motivated person such as these subjects can readily learn to discriminate with good accuracy between works considered by experts to be better or poorer when he can make these judgments and be informed immediately whether he is right; he cannot, however, be depended on to acquire such a discrimination simply from devoting an equal amount of time to studying the same pairs and judging them esthetically without being informed about the judgments of others. While this general contrast between the effects of experience with and without feedback is absolutely decisive, there remains the very intriguing suggestion that some people do increase their agreement with experts as a result of the no-feedback experience. This suggestive finding is very much worth following up in future research.

Training with feedback is hardly likely to receive educational application in the precise form in which it was used here. Such training would be more compatible with usual educational practices if it were accompanied by explanation and demonstration of some of the reasons why experts judge as they do for each pair. In the absence of specific test it seems likely that such enriched tuition would be more effective and more diversely interesting. The very bare kind of training employed here, however, did lead to a remarkably high level of discrimination. But what is more significant was revealed by some of the statements made in interviews: that even in its sparse, stripped form this training provided an introduction to visual art that was extremely interesting to some subjects and from which they extracted a very substantial degree of understanding of esthetic values.

2. Initial experience with feedback vs. experience with feedback following experience with no feedback.

As indicated in the earlier presentation of results, the effects of training with feedback were of the same general character regardless of whether this training had been preceded by judging the same pairs without feedback. In both cases correct prediction of expert opinion rose sharply during this training. But was there any significant quantitative difference? The results have not thus far been presented in a way to answer this question.

A partial answer may be provided by comparing the slopes which measure rate of learning shown by the two groups during training. The mean slope of the group which begins with feedback is steeper (+.20) than that of the group which has previously made judgments without feedback (+.15), and this difference is significant at the 5% level (t = 2.28, d. f. = 37). The level of performance attained does not, however, seem to differ as between the two groups; on the post-training test with easy pairs they differ by only one point (subjects with previous experience without feedback, mean of 91; subjects without this experience, mean of 92). The two groups are also nearly identical in their initial test means (59 and 61, in the same order). It appears likely, therefore, that the slower learning of the group with no-feedback experience is connected with their having learned something from this experience which gives them a beginning advantage in the training with feedback but no long-run advantage. A possible check on this may be obtained by considering the responses of the two groups on the very first magazine in the training with feedback. The subjects who had previously had no-feedback sessions predicted the experts' choice correctly on 17.3 of the 30 pairs, on the average, while the subjects without this prior experience did so on only 14.0 pairs. This difference is significant at the 1% level (t = 2.76, d. f. = 37). Here is additional evidence, then, of genuine gain from the non-feedback experience, and reason to attribute the slower learning of the non-feedback group to the fact that their gain was valuable at the start of the training with feedback but did not raise the level of ultimate attainment.

3. Easy pairs vs. difficult pairs.

Training in discrimination on the easy pairs was generalized to the difficult pairs. This is shown by the fact that the training with feedback led to an increase in score not only on the easy test but also on the difficult test. On the latter, one group showed an average increase of 13 points and the other group, 16. Only 2 out of 40 subjects failed to show an increase. We may thus conclude confidently that the distinctions between works of art in the difficult pairs have something in common with the distinctions represented in the easy pairs. But how about the effects of training with the difficult pairs?

No group was given initial training on the difficult pairs. We can reasonably expect that such training (with feedback, of course) would produce learning, but less learning than corresponding training with the easy pairs. What seems important to test is whether after the training with easy pairs, subsequent training with difficult pairs could lead to further improvement. Such improvement might especially be manifested in performance on difficult pairs to the extent that the distinctions represented there are of different sorts from those represented in the easy pairs. The results show that training on difficult pairs subsequent to training on easy pairs does not lead to improved performance. The mean slope of performance during the difficult training actually tends down instead of up, and changes in test performance from before to after difficult training are about equally divided between gains and losses, both on the easy test and on the difficult test.

The results do not, then, suggest that any distinctive educational value is to be found in the difficult pairs. The results are compatible with viewing the difficult pairs as sampling the same distinctions which characterize the easy pairs but representing them less regularly or dependably. The results presented in Chapter III, on personality correlates of esthetic judgment as measured with easy pairs and with difficult pairs, are also compatible with that view.

4. Low scorers vs. high scorers.

The 9 subjects who originally had high scores on the easy test of esthetic judgment and who went through one or more of the training procedures showed results that resembled in general nature those obtained with the 40 subjects who originally had low scores. Because of the very small number of persons involved, however, less confidence may be placed in the precise outcome. The high-scoring subject, too, showed no clear evidence of increased agreement with experts during a series of sessions without feedback. They showed overwhelming evidence of learning when feedback was provided for judgments on the easy pairs. When the training with feedback was given with difficult pairs, suggestive evidence appeared of a difference between high and low scorers; where the low scorers showed a barely significant trend toward poorer performance during this training, the high scorers showed a non-significant trend toward better performance. It is certainly reasonable to expect such a difference. The low scorers come to the difficult training with recently-acquired standards developed through their experience with the easy pairs. The high scorers come to it with this experience but also with pre-existing standards presumably developed in a variety of contexts. Because of recency, the standards of the low scorers might well be less stable in the face of challenge. And the standards of the low scorers seem more likely to be challenged because of their specificity of origin.

In the part of the experiment where everyone gains--training with feedback on the easy pairs--the high scorers gain at a lower rate (as measured in number of pairs judged correctly) since their performance is so high to start with. They remain above the low scorers in their final performance, however. On the post-training easy test 6 out of the 9 original high scorers have a higher score than the 104 out of 120 attained by the most successful learner among the original low scorers, and the other 3 are only a few points behind; the mean score of the two groups differs on the post-training test by 15 points (107 correct out of 120, as compared with 92 correct). On the post-training difficult test the separation of the two groups is still maintained, though it is less extreme; the means differ by only 8 points (81 vs. 73), and only 2 of the 9 original high scorers excel the score of 85 reached by one of the original low scorers.

The training with feedback on the easy pairs, then, offers excellent opportunity to learn a discrimination between works of art considered better and poorer by experts. The discrimination can under these conditions be improved by those subjects whose own initial esthetic judgments tends to correspond with that discrimination. The discrimination can also be acquired by subjects whose own initial esthetic judgments do not correspond with it. But the level of discrimination shown by these two groups remains different at the end of the training.

APPENDIX

INTRODUCTION TO APPENDIX

An incidental outcome of the research done under this contract is that a great deal of information was obtained about the personal preferences and esthetic judgments of male college students at Yale and Stanford. This information seems likely to be generalizable at least to include men undergraduates at other similar institutions, and probably more widely. The research was not directed at obtaining such information, which is not directly relevant to the main purposes of the inquiry, and there seemed no reason to treat it (at great additional expense) in the systematic and quantitative manner of the main inquiry. Yet this information seemed likely to be of great interest to college teachers of art and art history. It has therefore been worked up in the more impressionistic manner that available resources permitted. The resulting article, presented in this appendix, is in press in The Art Journal.

THE PRECONCEPTUAL EYE:

A Study of Preferences in Art Among College Men

Joseph Slate and Irvin L. Child

Preconceptions and prejudices seldom remain fixed—ideally, the mind expands, the eye evolves. But in dealing with students in fine arts courses or in contemporary art history, the artist and teacher assume this at some risk to communication. For although the language of art has changed radically since Manet, objections to its forms have not. The student may be unaware of it, but his preferences are in many cases historically as well as psychologically rooted. How to deal with him? A teacher can hardly know the aesthetic or psychological set of each student, but he can prepare to meet those objections that are common to many. Knowledge of such objections would be, we think, of intrinsic value in the teaching of art. It is with this in mind that we present here a set of findings about preferences of college men.

In 1959, 22 undergraduate men at Stanford University were asked to respond to 720 reproductions of paintings. The students, nearly all freshmen or sophomores, were chosen without respect to their interest in art. They were obtained, as paid subjects for psychological research, partly through soliciting in elementary psychology courses and partly through the student employment bureau. Nothing was said in advance about the project's concern with art, and, as might be expected, the group included students whose knowledge of art ranged from very high to negligible.

Students were seen individually, and each presented with sets of postcard-size color reproductions of paintings. Each set consisted of sixty highly diverse paintings, having in common only general subject matter. Categories included two sets each of paintings with groups of persons and religious paintings and one set each of landscapes, abstractions, still life, paintings rather arbitrarily labeled distorted or fantastic, paintings of women singly, men singly, children singly or in groups, and animals. These groupings were somewhat dictated by subject matter at hand, convenience for those who would be making judgments, and, more importantly, by the need for reducing the number of judgments that would be made simply because of such obvious subject matter preferences as for paintings of one sex over another, for paintings of animals over abstractions, etc. With a few exceptions, artists' names were not printed on the front of the cards, and students were urged not to turn over the cards for identification.

Students were asked to indicate relative liking for each painting in a set, from those least liked to those best liked, by sorting the sixty paintings out into ten piles of six each. Once Stanford students finished the project, a group average order of preference was calculated for each set of paintings. (Because of other purposes of the research, it was not a simple average that was obtained, but rather a factor score representing degree of popularity, based on a mathematical procedure known as factor analysis. But for present purposes this difference is unimportant.)

A year later, these procedures were repeated with a group of 22 Yale undergraduates whose services had been obtained in the same way. The average order of preference obtained from the Yale students was strikingly similar to that obtained from the Stanford students. (Just how similar, for the various sets, is one of the findings in a technical article which places this research in a context of psychological research on aesthetics. See I.L. Child, "Personal Preferences as a Measure of Aesthetic Sensitivity", Journal of Personality, 1962, Vol. 30, pp. 496-512). In some respects, too, this average order of preference throws light on the experience one of us has had in teaching art in still another college (Kenyon). This convergence of practical experience and research results from three widely separated colleges promises some fairly broad applicability to college men of the present day, and perhaps--in a lesser degree--to college women.

The most obvious agreement between the Yale and Stanford groups involved color. Paintings done in the cooler hues, of low saturation, and with an economy of hues, were preferred consistently throughout all categories. One might assume that a hostility to the use of pure color is associated with academic painting, and that perhaps these choices were made on that ground alone, for it is true that a high proportion of paintings favored were academic. This still does not explain the choice of cooler-hued, more simply hued abstract, i.e., non-objective paintings; relationship of hue to preference seemed, if anything, more consistent for abstract than for representational paintings. Preference for cooler hues, low saturation, and an economy of hues, can be seen to have an influence even within the work of a single artist as sampled in our study. When a late 19th or early 20th century painting was more highly preferred than others by the same artist, it was usually from his less coloristic period, such as Picasso's blue period.

There are exceptions to the hue preferences we have described; a high average rating was given some of the most widely publicized Van Goghs and Renoirs, despite their use of a variety of warm-hued and highly-saturated colors. It would appear, though, that such high-keyed colorists must have strong representational elements in their work to secure wide admiration by the college students of today. One immediately sees a hierarchy of preferences here; Gauguin, who proved unpopular with both the Yale and Stanford groups, was still preferred to such painters as Beckmann, Rouault, later Picasso, and Léger.

Color preference, however, seemed less important than accuracy in the representation of form.

The Chagall painting The Bride and Groom will serve as a good example of how antagonistic is the naive eye to distortion of form. His color in this reproduction is coolly hued and consequently should have been highly acceptable to this group. The colors also are conventionally placed: the bride wears white, the groom black, the sky is blue, the land is green. But Chagall is too close to that slightly askew world of children's art; he shatters the mirror, and consequently antagonizes those conditioned to representational work. His reproduction was ranked in the bottom ten in this set—57th out of 60 paintings—with only a Rouault, Seurat, and Léger following after it.

Léger and Rouault brought up another curious prejudice in these two groups.

Where the painter has conceded to some image, however stylized or abstracted it may be, students insisted that this image be set forth tonally. Paintings showing evidence of fine-line edging or heavy encasement of color areas in dark lines were highly unpopular. Lines also shatter the mirror; they exist neither in nature, nor in the academician's world, and both Léger and Rouault cared little for that world. Even among abstract paintings, those that trace form in heavy line were ranked low. There was no great objection to a calligraphic line as long as it did not frame areas of color; on the contrary, students seemed to like calligraphy for its own sake—a Mark Tobey, while not extremely popular, was in the upper 50 percent of the abstract group, and three Soulages paintings were ranked in the top ten in the abstract category by both Yale and Stanford.

This rule, like the others, has exceptions. Bernard Buffet, usually found in the upper 20 percent of the various rankings, encases his images in line; but he—if in a more limited way than Van Gogh, Renoir, and the early Picasso—has had his work highly publicized, and his distortions, relentless elongations, and subdued color did not move too radically outside the judges' experience. Except for Van Gogh, moreover, these artists have dealt in subject matter especially attractive to the young male: Buffet, with his wan garret self unbuttoned to the waist; Renoir and his young women softly endowed and inviting to the touch; and Picasso, in his classical period tenderly delineating themes of love and maternity. Even with such themes, the more linear work of Picasso or Buffet never was ranked highest. Renoir rarely introduced line for its own sake in his paintings, and his tonal impressionism was better liked by these students than the work of his contemporaries.

We should stress here that we were attempting to isolate factors that were sometimes associated and sometimes independent. What appeared as an exception to one rule, therefore, seemed in some instances to be clarified by another. As we have just mentioned, for example, subject matter sometimes cuts across prejudices against linearity or color. Two categories of work, abstract and fantastic, offered the advantage of reducing or even eliminating one of the major factors, that of subject matter. In these categories we were able to test various conclusions in the absence of subject matter, or else with subject matter transformed to such a degree that the students had to choose among various paintings which are, from the naive point of view, about equally distorted.

Oddly enough, there was high agreement on abstract painting, despite its great novelty to many of these students, and despite the fact that some bases of possible agreement were necessarily absent or reduced. The Yale and Stanford group averages pointed to exactly the same painting as the least liked, and the painting the Yale group favored most was ranked second by the Stanford group.

The abstract painting least liked is by Robert Delaunay, and that particular painter's work was ranked last by both Yale and Stanford in two different sets. Especially interesting to this inquiry is the fact that one of his disliked paintings is figurative and the other is abstract. The Delaunay abstraction, *Joie de Vivre*, is a target—to oversimplify—cut off at its extremities and inlaid with other smaller targets. The figurative work, *The Runners* (Staatsgalerie Stuttgart), is a flatly-patterned view of five men racing. What is common to both? Warm, brilliant hues, thickly and flatly applied, rough edges, a complete divorce from mirror reality, and both works bluntly brought to the forefront of the picture plane; the end runner's feet are cut off, and so are the target edges. This is objectionable to persons still oriented to the picture as a stage or window vista. Even in those abstract works preferred by the students, one sees this relationship, a preference for a dimensional quality, a certain distance between the observer and the thing observed. (But it also must be a logical distance, one must observe from a proper place. In the landscape category, seascapes looking to land, thus putting the viewer on the sea, were ranked low; but those landscapes with an open road leading from the middle of the canvas were popular. Students preferred open vistas easily entered into.)

Perhaps the need for staged or window-vista painting explains the reason for the obvious unpopularity of circular paintings and abstract work with circular elements in them. Delaunay's famous *Disks*, which is itself a circular painting, was ranked 51st of the 60 abstract reproductions, and all five of Kandinsky's later geometrical paintings—three strongly circular and two less so—were in the lower third. It may be, too, that the students simply objected to one-dimensional geometrical images which imply the use of drafting tools, for the lower third of the abstract paintings abounded in sharp-edged, line-bordered, rectangular and circular forms on single-hued backgrounds.

This use of single-hued backgrounds appeared to play some part in influencing student likes and dislikes. If a form is abstract, rather roughly put down, as in Soulages' work, and the form itself varies in color and weight so that it has some dimension, a single-hued background seems unobjectionable. However, forms that are extremely sharp-edged, bordered by a fine line, and laid over a single-hued background, hence giving the illusion of shallow dimension, are unpopular. Klee, Villon, Nicholson, Magnelli, and of course, Kandinsky, all produced such paintings at times and, in each such case, were ranked in the lower third of the abstract set. A single-hued background implies a divorce from environment, a flight from mirror reality that is unacceptable. We see this immediately in student judgments on representational work. Manet's *The Fifer*, the boy against the monochrome background, was very unpopular, as were the many Gothic, Byzantine, and Early Renaissance works that use a single-hued, usually gold, backing for a linear figure.

The use of gold coloring is of course a violation of the representational world, but one is uncertain how much the gold itself, and how much other features of the styles in which it is used, was responsible for the low rating given paintings that contain this color predominantly. Most of the paintings which employ gold leafing heavily are either of subjects or in styles unpopular with these students, e.g., the Virgin enthroned and the Annunciation and work done in the Gothic and Byzantine manner. It does seem that the popularity of Biblical and Christ-dominated scenes among the religious paintings reflects the Protestant ethos in America; scenes from the life of the Virgin were shunned, but the Nativity, if set in a natural environment, was popular. Sentimental portraits of Christ and bloodless crucifixions and depositions were preferred. Salvador Dali's *Last Supper* was among the most popular of the religious paintings, and one rather expected to see this preference for the comfortable super-real.

Two other Dali paintings in the category labeled fantastic were popular with both Stanford and Yale groups. In fact, one sees in this category some of the same prejudices operating; many of the most favored paintings are representational, even if their subject matter is bizarre. The most interesting preference, however, is for cubistic and futuristic paintings; they made up a high percentage of those in the upper third. Cubism then, one of contemporary art's earliest distortions of form, was relatively acceptable to this group as long as it was labeled "fantastic." One is reminded of the gallery cliché, "I'll accept it, as long as you don't call it art." Contrarily, students ranked very low the children's paintings in this category; fantastic or not, children are inherently incapable of mirror-image representation and shouldn't play the game.

There is another conclusion to be drawn from the fantastic and abstract groups. Those paintings that place in the center of the composition the dominant design element or the most appealing image, were more highly considered, although this appears less strong a determining factor than, say, color. The painting ranked first by Yale and second by Stanford in the abstract group has, as its dominant feature, a center that appears lit from behind by a light bulb. It is Bertini's *The Lamp of Thetis*, done in blues and blacks with striations of white at the center.

Centrality seemed particularly desired where the categories fall heavily to portraiture. The eye as a focusing machine, the eye as camera, was highly evident in those categories with subject matter frequently and popularly photographed. Consequently, if the painting was obviously meant to be a portrait, the judges preferred the portrait focused from the bust up; full-length portraits are not unconventional, and one must, after all, see the face clearly. But there were other, less visual, more amusing taboos. In portraits of men, an old man was preferable. (Most portraits are taken in maturity, and the judges always preferred the familiar.) Further, an old man fits the photographic cliché of the picturesque.

There is also a sexual factor involved in the categories of men and women; the judges—all young men—showed no distaste for paintings of young women finely gowned, but woe to the young man, primped and powdered or got up in a playful way. There seemed little prejudice against a man wearing the costume of his day; what tried the judges is what a man *made* of his day. Here, the utilitarianism of our society showed through most strongly; all harlequins and clowns, even if children, were relegated to the bottom half of the pile, and so was a monk playing a cello (by Corot, an artist otherwise invariably ranked high). Sober men, seriously engaged—including farmers and peasants, who are allowed a certain cheer—were popular. With respect to women the opposite appears true. Frivolous young things, elegantly-gowned artists, musicians, courtesans and other ladies clean but of questionable virtue were acceptable—if beautiful. Plain women went unappreciated; old women, even if Rembrandt's mother, were ranked low; and all women too madly hatted were suspect. Women were given more freedom in dress than were men, but there was a limit on how far a woman could go. Manet's *Victorine Meurend in the Costume of an Espada*, fell to the lower third of the category on women. If she were slender, beautiful, playfully engaged, she would have been more highly treated, but there she was and is, a middle-aged housewife, in a costume too tight, woodenly posing a cape and sword in the air.

There is an interesting fact we might add here. Both Yale and Stanford averages placed *the same ten paintings* at the top of the women's category. It seems that the eye of a young man in America is uniformly well-educated to beauty in women. It would be unwise to draw a conclusion as to the type of beauty preferred, as the sixty reproductions do not give that wide and divergent a range. Nevertheless, an oval face, large eyes, full lips, and a youthful plumpness are common to all of the favorites.

An artist or teacher of art may feel that all these bases of preference are irrelevant to student reactions to paintings *as art*. These students after all were deliberately encouraged to express preferences of a most personal, subjective, and variable nature. If they had been asked to make aesthetic judgments about the paintings, wouldn't the results be entirely different?

We think not. This conclusion is based on evidence from later research in which students were indeed asked to make aesthetic judgments. The material this time consisted of pairs of artifacts, rather than sets of 60, and the pairs were presented by projecting them on a screen simultaneously. The idea of using pairs in this way was suggested by the very stimulating work of the English aesthetician Margaret Bulley (*Art and Everyman, A Basis for Appreciation*; London, Batsford, 1951).

Each pair consisted of two works with some common basis of similarity, in style, subject matter, or both, but differing in aesthetic quality. The pairs, originally selected by graduate students in art, or others with similar background, were all subjected to the opinion of 14 judges—also graduate students in art—and only certain pairs were retained. These were pairs in which 12 or more of the judges on very brief inspection of the pair gave the same opinion as the original selector. Over a thousand pairs met this requirement.

From these pairs, 120 were chosen at random to make up a test of aesthetic judgment. Each pair was projected for about 15 seconds, and the student was asked to judge which was better aesthetically, that is, better as a work of art. (He was told that in case he could form no such opinion himself, he should then attempt to guess which was the experts' choice.) This test was given to 142 Yale undergraduates obtained as subjects for a psychological experiment. Again, no mention of art was made in connection with the solicitation of subjects. But since the research continued for over a year, it seems certain that some of the later students signed up because of an interest in art. Despite this fact, however, a look at the choices made by a sizable majority of the 142 indicates general agreement with the bases of preference registered in the postcard study.

The slide-project study amplifies the postcard study in another way. Because so large a number of students took part, it was possible to refine the inquiry. Of the 142 students, there were 52 who agreed with the experts on no more than 70 pairs out of the 120. These students grouped around the score of 60, which is what one might expect from a random basis of judgment.

What were the judgments of these 52 students like? Were they in fact random—judgments such as might result from tossing a coin? Not at all. There were some pairs on which they overwhelmingly agreed with the expert opinion, and other pairs on which they overwhelmingly disagreed. (Predictably, there were still more pairs on which they were nearly evenly divided, but these were much less frequent than if the judgments had been made at random.)

Here too the bases of judgment seemed to remain much the same. In pairs where both works of art were strange or distasteful to these students, they often chose work that held most closely to the visual preconceptions that we already have described. Of those 52 showing extremely poor agreement with the experts, only 30% held with the experts that a totally cubistic Picasso work was better than a rather faded reproduction of a semi-cubistic work with easily identifiable bottles and musical instruments. When confronted with two African masks, 84% chose the better one, but also the one with strong representational features. Of two painted Oceanic figures, the less brightly hued and the one having its anatomical parts more rightly situated, although abstracted, was judged better.

Distaste for linearity as evidenced in the other experiment held true here. A Rembrandt drawing in ink was shunned by all but seven per cent for a pencil drawing that allows for a certain amount of shading and tonality. An abstract blown-glass tree form was chosen by 86% over a similar one with linear curlicues inside it; and a calendar Swiss village won all but four per cent over a Dufy hill scene of flat washes and line.

However, when the choice was between a sketchy line drawing and primitive or children's art, the latter two always came out second best. If we were asked to suggest a work of art least considered art by undergraduate males we would safely put forth a pencil drawing by a child, and a child indifferent to the borders of his drawing pad. A sloppy edge after all chops away the illusion of a clean window vista. (Thus, John Marin's unpopularity. Marin leaves a border, but a ragged one, and his mixture of planes and lines is outside a cliché frame of reference. His vigor is no help to him either. An expressionistic approach to nature is so much an extension of a childlike intensity, so linear, and so insistent on fresh vision that the naive in art back away.)

In a further test with 19 of the 52 low scorers, there was greater verification of earlier tests. This time students had to make aesthetic judgments on 930 pairs of slides. Their preference amplified prejudices of the larger groups tested.

Mirror-image 19th century representation was preferred over all but impressionism as practiced by Renoir, an artist we have shown to be highly favored by these groups. Other impressionists seemed not to have fared as well; 89 per cent chose an illustrative and grimly-colored mother and child to a Mary Cassatt painting of the same subject. Impressionists' color also bothered this group; 84 per cent judged Monet's more coldly hued painting of haystacks better than one in the same series caught vividly at sundown. However, impressionism was preferred over distortion or highly-stylized advertising art. Pen or ink drawing that lacks modeling was rarely judged better than its opposite; of two Picasso minotaur drawings, the one that approaches a kind of tonality was chosen by 79 per cent. Judgments indicated a preference for dramatic naturalism; thus the Goya execution (The Shootings of May Third) was chosen by 89 per cent over Manet's The Execution of Maximilian with its stiff, frontal formality. Dislike for shallow dimension, particularly where forms are set before a flat-hued ground, was amplified. Children's work fell greatly into disrepute: a linear Klee or heavy-lined Rouault, both highly unpopular, were chosen over or paired indecisively against a painting by a child.

There was greater verification in this group for indications that male students disliked circular elements in abstract work. When paintings with circular elements were paired against those without, 80 per cent or more chose those without circles. Evidence bore out earlier conclusions that round paintings or oddly-shaped paintings were shunned, even when students favored the works' content. When circular elements were paired, students split indecisively. The one exception to this involved a pairing of two paintings with circular elements, one of which was crudely done, as by a child, multi-colored, and with its blotted forms on a white ground. In this case, there were so many prejudicial elements in one as opposed to the other that 89 per cent of the students marked preference for what they considered the less objectionable painting, although it contained highly unpopular clean-edged circular forms.

The fact that students tested were males undoubtedly accounted for two cases of unanimous dissent from the judges. These both involved beautiful women. A 19th century Russian work, Girl With A Mortar, by Venitsianov, was chosen over Vermeer's Young Girl With a Flute, and one can only assume that the round-faced farm girl is so potently attractive that Vermeer's Chinese-hatted worldling simply had not a chance. Easier to understand is the student preference for Titian's Mary Magdalene over El Greco's painting of the same subject. El Greco's elongations were accepted when painted in a religious context, but he was in trouble when none was attached. (These pairs were not labeled nor were the Magdalenes marked with halo.) Over and over, students have rejected either heavily-stylized work or paintings highly manneristic.

Students agreed with the experts unanimously far oftener than they disagreed unanimously; but in most cases these works were commonly known and accepted. One finds that all students tested preferred Gainsborough's Blue Boy and Whistler's mother (Arrangement in Grey and Black No. 1), both paired against inferior work.

That familiarity might be a source of aesthetic sensitivity was suggested by the students' reaction to architecture, an art form with which everyone necessarily has much experience. Of the 930 pairs of slides, 18 pairs were of architectural structures, as follows: two pairs on houses, two on bridges, three on church entryways, two on three-story structures, four on skyscrapers, one on Indian temples, one on Greek-derived structures, one on cathedrals, and two on New England churches. All 18 structures selected as better by the students were the ones the judges considered as better; 58 per cent agreed with the judges on one three-story structure, and this was the lowest percentage of agreement. The others ranged in agreement from 74% up to 100% with houses and skyscrapers gaining the highest percentage of agreement. Obviously then, students were knowledgeable about structures within their visual experience, and they seemed easily capable of transposing this knowledge to similar forms. Their appreciation of the need for balancing masses on skyscrapers may have assisted their judgment on the cathedrals, churches, and Indian temples.

We will not venture to press this argument with much vigor on such slight evidence, as we can readily see objections. Simplicity and functionalism in architectural forms may be the common preference in America. A silo or the Golden Gate Bridge is as commonly known to the naive as September Morn, but where the former informs his taste the latter debases it. The public preference may happen to coincide with the experts' judgment for reasons we might well call "historical accident."

While we have enjoyed speculating between ourselves about the possible sources of various preferences we have described, we hesitate to read more into this study. How commonly valid are these findings remains uncertain. We claim only to have inquired into the preferences in art of men, college students, at Yale and Stanford (with some observations at Kenyon). Nevertheless we have been surprised at the similarity between student objections to art forms and the objections voiced by critics when Manet first began urging his fellows into the 20th century. That was a hundred years ago, but such preconceptions appear to have changed little since.